The Economics of Preference Heterogeneity

ISBN / EAN: 978-90-9026000-6

Cover Design: Patrick Ketelaars
Print: Ipskamp Drukkers, Enschede

The Economics of Preference Heterogeneity

Een wetenschappelijke proeve op het gebied van de Managementwetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 18 april 2011
om 13.30 uur precies

door

André Antonius Johannes van Hoorn
geboren op 13 augustus 1976
te Helmond

Promotor: Prof. dr. E.-M. Sent

Copromotor: Prof. dr. S. Beugelsdijk, University of Groningen

Manuscriptcommissie:
Prof. dr. E. de Jong
Dr. C. Bjørnskov, Aarhus University
Prof. dr. H.J. Garretsen, University of Groningen

Contents

Part I

An Overview and Evaluation of Empirical Approaches to Measuring Preference Heterogeneity

Part II

Subjective Well-Being and (Heterogeneous) Experienced Preferences

Part III

Assessments of Heterogeneity in Experienced Preferences

Part IV

Conclusion

List of Tables

List of Figures

Preface

How to start the preface of your dissertation (not least when you realize it will probably end up being the most read part)? I have heard more than one story about people struggling for a long time with just their opening sentence. As is, I have already passed this obstacle…

To get down to the business at hand, let me commence with a brief description of how this thesis came about. I started working on my PhD thesis in August 2005 when I became a Junior Researcher at the Department of Economics of Radboud University Nijmegen. Throughout my time as a Junior Researcher, I have been drawn to many interesting topics; indeed, for me, one of the biggest pleasures of being an academic and doing a PhD is that it allows you to constantly uncover new, exciting research questions and work on them. In this light, I am unable to pinpoint any defining moments in my PhD project. I remember settling on the specific topic for my thesis around June 2008. Similarly, I recall that about two years later, in July 2010, the pieces really fell together, meaning I had an almost completed draft version. All the rest is history I am tempted to say.

Along the way I did receive a lot of help and support, however, and surely a preface to a (any?) dissertation would seriously disappoint its readers if it did not include some (grateful) acknowledgements. Without further ado, I want to take this opportunity to thank my supervisors, Esther-Mirjam Sent and Sjoerd Beugelsdijk. It is with great pleasure that I go on record saying how much I appreciate their involvement in the present thesis and how thankful I am for the opportunity and freedom they gave me to pursue my broad research interests. The book before you would be much less complete if it were not for their valuable comments and suggestions.

Sjoerd and I are actually colleagues again at the University of Groningen and so is the third person whose contribution deserves very grateful acknowledgement, Robbert Maseland. My gratitude goes beyond the co-authored chapter and excerpts of joint work included in this dissertation. It extends to all the work we have engaged in outside the PhD project and to the many discussions that we have had about possibly interesting research questions. Somewhat paradoxically, the latter often left me understanding the world a bit better, but also realizing how complex it really is.

Less directly related to either the present dissertation or on-going research, my (former) colleagues in Nijmegen are also well-worth special mention. Here I want to thread

carefully. I am thankful to many people, but I do want to signal out some members of the PIJEN group: Roger, Ellen, Remco, Ferdy, Ramzi, Floris, Sandor, Aloys, and Aagje, to name but a few of the illustrious names associated with the group that contributed enormously to much of the joy I experienced during my time in Nijmegen.

Whilst I was associated with Radboud University Nijmegen, I have further enjoyed the hospitality and warm atmosphere of the Statistics Directorate of the OECD in Paris and the World Database of Happiness at Erasmus University Rotterdam. I would like to thank Ruut Veenhoven and Jon Hall, my project manager at the OECD, for offering me the opportunity to visit them for an extended period. I am also very grateful to all the colleagues at these two institutes for providing me with a warm and inspiring environment.

Over a year ago, I left Nijmegen for Groningen. Numerous people have not only made the transition to the University of Groningen a smooth one, but have also made me feel very welcome there. More generally, they make the fifth floor of the Duisenberg building a very nice place to be. I look forward to further pleasant times to come.

Finally, special thanks go to my family and friends. Though they may not realize it, they have contributed a great deal to this thesis. No doubt, they have helped make me the person I am today, but they have also offered me much in terms of love, friendship, and support. I want to avoid mentioning names, as I am thankful to all of you, but let me emphasize how fortunate I feel with you as a family or with having had our paths cross at some point in time.

André van Hoorn
Groningen, February 2011

Every man is in certain respects:

a. like all other men,

b. like some other men,

c. like no other man.

- Kluckhohn and Murray (1948: 35)

Chapter 1

Introducing: The Economics of Preference Heterogeneity

1.1. STUDYING HETEROGENEOUS PREFERENCES

In 1997 a group of economists, anthropologists and other social scientists formally joined forces in The Norms and Preferences Network.[1] The network was founded as a MacArthur Research Network and sought to bring together people with different backgrounds for the interdisciplinary study of preferences. Amongst its members were George Akerlof, Samuel Bowles, Colin Camerer, Ernst Fehr, Edward Glaeser, Avner Greif, Daniel Kahneman, David Laibson, George Loewenstein, Casey Mulligan and Paul Romer. It was chaired by anthropologist Robert Boyd and economist Herbert Gintis.

The most prominent research to come out of The Norms and Preferences Network is the Foundations of Human Sociality project led by anthropologist Joseph Henrich (Henrich et al. 2004). Building on early work by Henrich (2000), this project used dictator and ultimatum games to measure inequity aversion or the preference for a fair distribution in several small-scale societies (Henrich et al. 2004; Henrich et al. 2005; Henrich et al. 2006). In line with earlier findings (Henrich 2000; Henrich et al. 2001), results showed considerable heterogeneity in the taste for fairness amongst these societies and between people from these societies and a sample of students at the University of California Los Angeles. The project was funded by The Norms and Preferences Network and several of its members, including Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis, acted as co-investigators.[2]

Other work coupled with the network has been less empirical, and more conceptual/theoretical. Specifically, George Akerlof has over the last decade written extensively about the incorporation of identity—people's notion of who they are—in economics. Most of this work has been joint with Rachel Kranton and argues that preferences or tastes partly derive from identity and social norms on how one should behave (Akerlof and Kranton 2000, 2002, 2005, 2008, 2009; Akerlof and Shiller 2009). Based on this work, his 2007 presidential address to the American Economic Association makes a plea for macroeconomists to start looking at people's motives, arguing that a better understanding of identity and accompanying tastes generates insightful new perspectives regarding, for

[1] The information in this paragraph is taken from the website of The Preferences Network or The Preferences Project as it also has been called, http://www.umass.edu/preferen and of the John D. and Catherine T. MacArthur Foundation, http://www.macfound.org/site/c.lkLXJ8MQKrH/b.951815/k.1CE0/Research_Networks__The_Nature_and_Origin_of_Preferences.htm. Gintis and Romer (1998) discuss research goals for the network.

[2] See http://www.hss.caltech.edu/~jensming/roots-of-sociality for more information about this project and the follow-up project named the Roots of Human Sociality (e.g. Henrich et al. 2010).

example, the independence between consumption and current income and between long-run inflation and unemployment (Akerlof 2007; see, also, Akerlof 2002).

This thesis is on the study of preferences, specifically heterogeneity therein. Just like the Norms and Preferences Network, it fits the wider trend that is occurring in economics, where researchers are increasingly studying the shape and content of the utility function and, particularly, how these differ among groups or classes of people. A string of recent research examines, for example, religious variation in preferences, particularly those towards work (Arruñada 2010; Guiso et al. 2003, 2006; McCleary and Barro 2006; Schaltegger and Torgler 2010; Torgler and Schaltegger 2009), or the effect of growing up under Communism on people's preferences (Alesina and Fuchs-Schündeln 2007; Blanchflower and Freeman 1997; Corneo and Grüner 2002; Shiller et al. 1991, 1992). Other work analyzes specifically how social norms affect the preference for work, finding that lower peer unemployment increases the loss of well-being experienced during an unemployment spell (Clark 2003; Clark and Oswald 1994; Shields and Wheatley Price 2005; Stutzer and Lalive 2004). Using the same basic approach, Alesina et al. (2004) and Di Tella and MacCulloch (2005) assess partisan differences in preferences, finding that the left and the right differ in the happiness loss they suffer from inequality, inflation, and aggregate unemployment. Some further contributions bring in quantitative indices of cultural differences in preferences developed in neighboring social science disciplines such as those of Hofstede (1980, 2001). An example is the application of Hofstede's (1980, 2001) measures in empirical analyses of cross-country patterns in entrepreneurial activity (Freytag and Thurik 2007; Wennekers et al. 2007), openness to trade (de Jong et al. 2006) and economic performance (Gorodnichenko and Roland 2010), or topics in finance such as trading strategies, portfolio composition and financial development (Beckmann et al. 2008; Beugelsdijk and Frijns 2010; Chui et al. 2010; Fidrmuc and Jacob 2010; Garretsen et al. 2004). Finally, there is a somewhat longer tradition of using economic games like dictator, ultimatum and public good games, as well as other types of choice experiments, to measure gender and country heterogeneity in preferences for fairness, risk aversion, and willingness to contribute to public goods. As with the project on the Foundations of Human Sociality, this body of research finds substantial differences in preferences between men and women (Andersen, Bulte et al. 2008; Andersen, Harrison et al. 2008a, 2010; Andreoni and Vesterlund 2001; Borghans et al. 2009; Croson and Buchan 1999; Gneezy et al. 2009; Holt and Laury 2002; Schubert et al. 1999), and between different (industrialized) countries (Burlando and Hey 1997; Chuah et al. 2007, 2009; Croson and

Buchan 1999; Gächter and Herrmann 2009; Gächter et al. 2010; Herrmann et al. 2008; Roth et al. 1991; Weimann 1994).

In the literature studying heterogeneous preferences different approaches to giving empirical content to preference heterogeneity can be discerned. This thesis categorizes and evaluates these approaches, finding a particularly promising approach, which we develop further and apply to the study of some issues that are long-standing and/or on which the existing evidence is particularly mixed. An outline detailing the content of this thesis, and the main questions it takes up, follows.

As a further introduction to the study of preference heterogeneity, below I first give some brief background of heterogeneous preferences in economics (Section 1.2). In addition, I devote a small part of this introductory chapter to sketching a picture of the context in which a distinct literature on the economics of preference heterogeneity appears to be developing (Section 1.3). The detailed overview of the thesis then concludes.[3]

1.2. DE GUSTIBUS (NON) EST DISPUTANDUM

The interest in tastes or preferences described in the first section of this introductory chapter, measuring them and assessing heterogeneity across groups (societies, individuals with different religious backgrounds, et cetera), is not traditional to economics (cf. Arrow in Swedberg 1990: 148; Gintis and Romer 1998: 2). In fact, the opposite rings more true (cf. Gintis and Romer 1998: 2). Scitovsky (1976: xii), for instance, finds that "[p]eople's tastes, how they spend their money and arrange their lives, are matters economists have always regarded as something they should observe, but must not poke their noses into."

The classic argument against heterogeneous and in favor of homogeneous preferences in economics is made by Stigler and Becker (1977) in their famous "de gustibus" paper. In it, they argue that economic phenomena should be analyzed strictly in terms of prices and constraints (see also Becker 1976, 1996 and Michael and Becker 1973). As they put it, "tastes neither change capriciously nor differ importantly between people." And "[o]n this interpretation one does not argue over tastes for the same reason that one does not argue over

[3] To be complete, apart from this introductory chapter and the concluding chapter (Chapter 8), the chapters have been written in plural active style. This is done to improve readability, but also reflects that two of the following chapters (3 and 7) are based on joint work.

the Rocky Mountains—both are there, will be there next year, too, and are the same to all men." (Stigler and Becker 1977: 77).

Formally endowing preferences the properties Stigler and Becker (1977) ascribe to them implies simply that any consumption bundle z yields the same utility to all individuals j, $u_j(z) = u(z)$, and that preferences do not change over time ($\partial u(z)/\partial t = 0$ while keeping consumption constant, $\partial z/\partial t = 0$).[4] The methodological insight that is expressed in Stigler and Becker's (1977) assumption of preferences that are constant and identical goes much deeper, however. In short, the idea is that invoking heterogeneous (and/or non-constant) preferences is ad hoc, meaning it can explain anything and hence nothing. More specifically, when choices reveal preferences, a preference-based account of economic behavior and outcomes amounts to a tautological argument. The shunning of preference heterogeneity disciplines researchers—and for Stigler and Becker economic analysis ends when preferences enter the argument. At that point, the issue at hand is passed on "to whoever studies and explains tastes (psychologists? anthropologists? phrenologists? sociobiologists?)" (Stigler and Becker 1977: 76).[5]

The turn towards the study of preference heterogeneity in economics constitutes a clear break away from the Stigler-Becker tradition. In no way does it constitute a slackening of methodological discipline, however. Rather, it derives from improvements in measurement methodologies, catching up with Stigler and Becker's (1977) de gustibus argument and enabling the inclusion of heterogeneous preferences in economic analyses, next to the

[4] Other properties of the utility function are, of course, that utility increases with consumption but at a diminishing rate: $\partial u/\partial z > 0$ and $\partial^2 u/\partial z^2 < 0$.

[5] Stigler and Becker (1977) offer an alternative approach, which takes the household production function as the relevant unit of analysis. These functions are used to model the transformation of goods and services into commodities, which are the ultimate carriers of utility. Production functions can change and be non-identical for different individuals, whilst preferences remain constant and homogeneous. This proposed framework has been criticized as equally vulnerable to ad hoc assumptions (e.g. Cowen 1989). Herbert Simon (1981: 58) conveys this critique most strongly where he argues that "[i]t may be doubted whether anything is gained by trying to rescue the traditional view of utility with such heroic measures. If, to continue the example, we do not wish to speak of a change in utility function as the result of listening to music, then we must postulate within the human head a production function (itself changeable by experience) that manufactures musical pleasure from musical listening. We have merely relocated "taste" from the utility function to that hypothetical new production function. It would seem more parsimonious simply to regard the utility function as an evolving structure." In supplement to these arguments, we may note that, empirically, it is often not possible to distinguish differences and changes in the household production function from genuinely heterogeneous and non-constant preferences.

traditional elements of prices and constraints. Underlying Stigler and Becker's (1977) case for homogenous preferences is the premise that preferences cannot be observed (except, as mentioned above, in the form of the very same economic outcomes that preference heterogeneity is supposed to illuminate). Over the last 35 years, a lot of progress has been made on this count, however. The increasing ability to give empirical content to heterogeneous preferences makes it possible to use preference heterogeneity in economic analyses in a way that is far from tautological or ad hoc but, in fact, strongly based in empirical observation (cf. Caplan 2003). As a result, the Stigler-Becker argument for taking preferences as constant and identical has lost part of its scientific rationale.

In line with Stigler and Becker's (1977) methodological insights, measurement occupies a central place in the emerging economics of heterogeneous preferences. In fact, a key part of this new literature is devoted to pushing forward the measurement of preference heterogeneity, developing ever more meaningful and comprehensive measures of how and to what extent preferences differ across groups.

The importance of measurement (and developments therein) also pops up in the discussion of mainstays for the increasing interest of economists in studying preference heterogeneity to which I turn next.

1.3. MAINSTAYS FOR THE RISE OF AN ECONOMICS OF PREFERENCE HETEROGENEITY

1.3.1. New Institutional Economics: An Important Intellectual Backdrop

Although economics until recently appears to have shunned the idea of preference heterogeneity, the possible importance of non-homogeneous preferences for economic analyses has not gone unrecognized. Particularly, new institutional economics (e.g. Greif 1994, 2006; North 1981, 1990, 2005; Williamson 1985, 2000; see Alston 2008 for an overview) has rather consistently argued for a perspective on economic phenomena and outcomes that takes explicit account of such heterogeneity. This field finds that since formal institutions are placed in specific historical and social contexts, the economics of these institutions cannot be studied without paying attention to societies' values or culture—that is the informal institutions in which formal institutions are embedded (Greif 2006; North 1990,

2005; Williamson 2000).[6] Williamson (2000: 596-600) thus identifies four levels of social analysis, putting embeddedness at the top. Cultural values or norms (Level 1) shape the institutional environment (polity, judiciary, bureaucracy; Level 2). These formal rules of the game then go on to affect governance (Level 3) and, ultimately, the allocation of resources (Level 4).

In Williamson's (2000: 597) framework, neoclassical economics is concerned with resource allocation (Level 4), whilst new institutional economics concerns itself mostly with questions concerning governance and institutional environment (Levels 2 and 3); which institutions or governance structures, for example, are most conducive to economic prosperity, and how do they come about? The economics of preference heterogeneity ventures into the first level of analysis (Level 1), that of values and culture.

In economics, such authors as Lal (1998) and Landes (1998) have also traversed this path in their historical case studies of how cultural values and other systematic differences in preferences shape economic, social and political life (see, also, Clark 2007; Greif 1994, 2006, Fukuyama 1995, Harrison 1992, Harrison and Huntington 2000, Mokyr 2008, and Putnam 1993). Another approach is quantitative and revolves around empirically studying cultural differences such as variation in values or preferences for different groups of people, and also relating these to economic outcomes. This literature is rapidly growing in economics; in addition to the studies mentioned earlier, see, for example, Antecol (2000, 2001), Blau et al. (2011), Fernández (2007), Fernández and Fogli (2006, 2009), Finkelstein et al. (2008, 2009), Ichino and Maggi (2000), Lopez and Lozano (2009), Phelps (2006), Phelps and Zoega (2009), Tabellini (2008, 2010), Williamson (2009), and Williamson and Mathers (2010).

Next to measurement, work such as the above on heterogeneous preferences as a possible source of economic differences constitutes a second key part of the economics of preference heterogeneity literature. Note, however, that for quantitative studies like the ones just mentioned measurement remains crucial—this holds independently of Stigler and Becker's (1977) methodological insights.

Shifting attention back to measurement we can see the emergence of an economics of preference heterogeneity running parallel with developments both outside and within economics. I expand on this below.

---

[6] I explicitly mention the embeddedness of formal institutions in informal institutions as the concept of embeddedness is widely used to convey the idea that social and cultural context matter (see Williamson 2000: 596-597 for an example). Let me also note, however, that the concept has been criticized, not least on the grounds of lacking a precise definition (e.g. Hodgson 2008: 139).

1.3.2. Cross-Cultural Psychology: An Empirical Pillar

Outside economics, disciplines such as cross-cultural psychology, and also sociology, have historically devoted much effort to measuring such entities as cultural values and national character (cf. Stigler and Becker 1977: 76). An important pioneer is Milton Rokeach (1973) who came up with an inventory of human values that has received much follow-up.[7] Another well-known researcher of individual preferences and attitudinal traits is Shalom Schwartz (1992; Schwartz and Bilsky 1987, 1990) who argues that human values break down in seven categories that are universally shared. The most popular work undoubtedly is that of Geert Hofstede (1980, 2001). His framework of cultural differences between countries along five dimensions is among the most cited works in social science (Bond 2002: 73; Hofstede 2001: xvii). Other aggregate-level measurement efforts are the GLOBE (Global Leadership and Organizational Behavior Effectiveness) project which overlaps with Hofstede's (1980, 2001) framework but extends it to nine dimensions (House et al. 2004), and Schwartz's translation of universal human values to country indicators (e.g. Schwartz 1994; Schwartz and Sagiv 1995). Finally, the World Values Survey (WVS), running since 1981 and currently coordinated by Ronald Inglehart (e.g. 1990, 1997; Inglehart and Baker 2000), is a very rich source containing data on individuals' values and attitudes concerning a wide range of issues. The entire WVS dataset is publicly available (European Values Study Group and World Values Survey Association 2006; World Values Survey Association 2009), and scores can be aggregated and analyzed at different levels. This contrasts with the indicators of Hofstede (1980, 2001) or GLOBE which are available at the country level only.

Data and findings from these large-scale measurement exercises provide the means to move beyond the assumption of homogeneous preferences and study the embeddedness of economic activity and formal institutions. Indeed, as alluded to above, economists increasingly use them for exactly this purpose (e.g. Beckmann et al. 2008; Chuah et al. 2007, 2009; Chui et al. 2010; Freytag and Thurik 2007; Gächter et al. 2010; Guiso et al. 2006; Herrmann et al. 2008; Tabellini 2008, 2010; Wennekers et al. 2007; Williamson 2009), in this way following related disciplines like political science, management and international business (IB), in which empirical work commonly uses the WVS or the GLOBE and Hofstede (1980, 2001) frameworks (e.g. Franke et al. 1991; Granato et al. 1996; Husted 1999; Inglehart

[7] Some earlier work is by Allport and Vernon (1931), Kluckhohn and Strodtbeck (1961), and Morris (1956).

1997; Jackman and Miller 1996; Kogut and Singh 1988; Park and Ungson 1997; Swank 1996).[8]

With this thesis I contribute to this empirical literature, whereby the focus similarly is on systematic differences in preferences between groups, such as individuals with different religious backgrounds. We should note, however, that in principle it is also possible to conduct small-scale studies of the economics of preference heterogeneity at the level of the individual. This would involve measures that are so finely grained that they are able to capture each individual's preferences in a unique score. Possible means to do so are provided by personality psychology, which has developed measures of individual traits—almost without exceptions based on survey items—that can proxy for differences in preferences (cf. Borghans et al. 2008: 975; Caplan 2003). The popularity of measures of personality traits in economics research is growing (see Borghans et al. 2008 for a survey). Given that the study of economic and social phenomena always involves some form of aggregation, I do not pay much attention to the possibility of using measures that provide unique scores for all individuals. Some of this thesis' findings and conclusions concerning the study of the economics of preference heterogeneity equally apply to the personality literature in economics, however.

1.3.3. Behavioral Economics: A Source of Method and Substantive Interest

In addition to its foundation in cross-cultural psychology, the rise of the study of preference heterogeneity in economics resonates with the wider trend known as behavioral economics (cf. Camerer 1999, 2003; Camerer and Loewenstein 2003; Heukelom 2009a; Mullainathan and Thaler 2000; Rabin 1998, 2002; Sent 2004; see, also, Akerlof and Kranton 2009, Chapter 4). Behavioral economics aims to increase the psychological realism of economic theory and thereby enhancing its explanatory power (e.g. Camerer and Loewenstein 2003; Rabin 2002). More specifically, "the goal is to find parsimonious utility functions, supported by psychological intuition" (Camerer 2003: 101; see also Simon 1987). For this purpose behavioral economics takes insights and results from psychology (again), which include empirical evidence on the heuristics people use when making judgments, how they discount future outcomes, and how they weigh gains and losses. In addition, behavioral economics applies techniques and methods from psychology to study these issues further.

[8] Beugelsdijk and Maseland (2011) and de Jong (2009) review much of the literature in economics and IB.

Concerning the latter, psychologists often use behavioral experiments. Behavioral economists on their part take these well-established experimental methods, adapting them, and then using them to test economic theories and to open up the concept of Homo Economicus to critical scrutiny.[9] Experimental results show that, for example, people do not have strictly self-regarding preferences as the canonical model has it, but have other-regarding preferences as well. To go from these studies of the shape and content of the utility function to assessments of systematic variation in preferences such as between the sexes (e.g. Andersen, Bulte et al. 2008; Andersen, Harrison et al. 2008a, 2010; Andreoni and Vesterlund 2001; Bolton and Katok 1995; Borghans et al. 2009; Eckel and Grossman 1996, 1998, 2001; Gneezy et al. 2009; Holt and Laury 2002; Jianakoplos and Bernasek 1998; Powell and Ansic 1997; Schubert et al. 1999; Solnick 2001) requires only a small step. In this kind of work, preference heterogeneity—Level 1 in Williamson's (2000: 597) framework—is linked to resource allocation—Level 4 in the same framework—directly. An important goal of assessments of gender differences in preferences, for example, is to improve our understanding of the different labor market outcomes for men and women (Croson and Gneezy 2009; Eckel 2008).

1.3.4. Conclusion: The Study of Preference Heterogeneity in Economics

The above discussion finds that the study of preference heterogeneity has different foundations in established fields, both within and outside economics. We can view institutional economics as providing a rationale for studying preference heterogeneity and cross-cultural psychology as providing the empirical means to undertake such research. Behavioral economics has a substantive interest in preferences but also deals with method and measurement. Psychological realism is an important issue of substance: the assumption of heterogeneous preferences is more realistic than that of homogeneous preferences (cf. Kluckhohn and Murray 1948: 35). The use of psychological methods on the other hand is an important factor enabling the study of preference heterogeneity. From this perspective, if the study of the economics of preference heterogeneity must be classified and placed somewhere, it is probably best viewed as a subfield in behavioral economics.

[9] As I discuss in more detail in Chapter 2, a major difference in the economists' and psychologists' application of experiments is the former's extensive use of (often substantial) monetary payoffs.

Beyond that, behavioral economics appears to have made the use of psychological data and methods more accepted in economic analyses. Behavioral economics therefore not only provides an interest in preference heterogeneity and the means to study it, but it has also paved the way for researchers to break with the Stigler-Becker tradition of relegating the study of preferences to other disciplines. In fact, the disciplinary scope of studying the economics of preference heterogeneity reaches well beyond the economics discipline. Psychology's historical concern with measuring attitudes, cultural values, et cetera (e.g. Rokeach 1973; Hofstede 1980) and the extensive use of measures of preference heterogeneity in quantitative political science, management and IB research (e.g. Franke et al. 1991; Granato et al. 1996; Husted 1999; Inglehart 1997; Jackman and Miller 1996; Kogut and Singh 1988; Park and Ungson 1997; Swank 1996) serve to highlight this point.

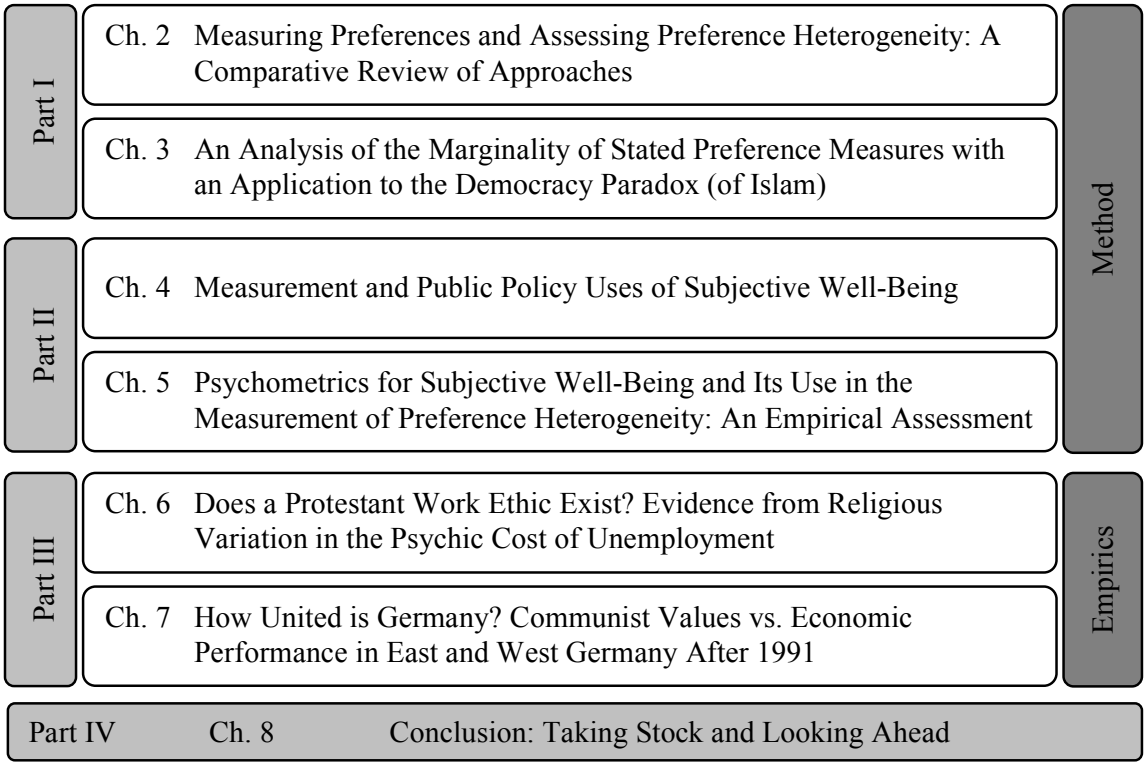
In this thesis I try to do justice to the cross-disciplinary nature of the study of preference heterogeneity as much as possible. Whilst the approach in this thesis is very much founded in economics, I regularly link to neighboring social science disciplines. This is reflected most clearly in the specific issues within the economics of preference heterogeneity that this thesis takes up. I have taken care to ensure that the heterogeneity in the preferences studied in the empirical chapters is not only of interest to economists but also to other social scientists, sometimes as part of long-standing research traditions.

1.4. THESIS OUTLINE

This thesis adds to the literature on the economics of preference heterogeneity in different ways. Its primary contribution is that it pushes forward a recently developed approach to assessing preference heterogeneity. The study of the economics of preference heterogeneity relies on three different methods of measuring how preferences differ across groups, revealed preferences (RP), stated preferences (SP), and what can best be dubbed "experienced preferences" (EP). As we would expect, these are the same methods that can be discerned in the measurement of preferences more generally, as for the purpose of valuing public goods or constructing money-metric welfare indicators (e.g. Fleurbaey 2009: 1054; Frey et al. 2007, 2010; Stutzer and Frey 2010: 23). Of these three methods, the RP and SP approach are fairly standard and well-known, although there have been interesting advances in the former, as I discuss in more detail in Chapter 2. The EP approach on the other hand is less well-known as it has been developed only recently.

The EP approach directly assesses the shape and content of the utility function. It uses measures of self-reported happiness or satisfaction (i.e. subjective well-being), which proxy for so-called experienced utility (e.g. Alesina et al. 2004; Di Tella and MacCulloch 2005; Frey and Stutzer 2002; Frey et al. 2009, 2010; Helliwell and Barrington-Leigh 2010a; Rabin 1998; van Hoorn et al. 2010). Experienced utility can be seen as "the hedonic quality" of an outcome, and must be distinguished from decision utility, the "weight of an outcome in a decision", which is the traditional utility concept in economics (Kahneman et al. 1997: 375). The EP approach to measuring preference heterogeneity then involves estimating heterogeneous *experienced* utility functions, where the coefficients of the determinants of happiness or satisfaction are allowed to vary with social norms (Clark 2003; Clark and Oswald 1994; Shields and Wheatley Price 2005; Stutzer and Lalive 2004) or with partisan identification (Alesina et al. 2004; Di Tella and MacCulloch 2005; Di Tella et al. 2010) to mention two examples.

Figure 1.1, Overview of the Thesis.



The thesis consists of eight chapters—including this introductory chapter—and four separate parts. Figure 1.1 gives a schematic overview. As a prelude, here I first give a brief

outlook on the contents of the thesis. A more detailed discussion, which also expands on the main issues taken up in the different parts and chapters, follows.

Part I of this thesis compares and evaluates the RP, SP and EP approaches to assessing preference heterogeneity, also elaborating their underlying measurement methodologies. The main aim of this review is to identify key differences (and some similarities) between the three approaches as applied in the study of the economics of preference heterogeneity. I thereby focus on two properties of the approaches and measures particularly relevant for their application in the study of the economics of preference heterogeneity, namely (i) psychometric quality (reliability and validity) and (ii) "usability." Results indicate that the EP approach has some potential well worth delving into more deeply. Part II of this thesis therefore presents a very detailed examination of this approach, focusing on reliability and validity issues arising in the context of using subjective well-being (commonly abbreviated as SWB) to measure preferences and heterogeneity therein. I thereby first go down to the nitty-gritty of the SWB construct—how it is defined, how it is measured, and how it can be applied—before presenting empirical evidence on the psychometric quality of SWB indicators and EP measures of preference heterogeneity. The assessment of the psychometric quality of EP measures is the very first of its kind and my filling of this gap in the literature contributes to the development of the EP method. In addition, the psychometric assessment of EP measures provides the foundation for the application of the approach to study specific open issues in the literature on the economics of preference heterogeneity in Part III of this thesis. Applying the EP method I am able to shed new light on existing (and long-standing) questions about specific sources of preference heterogeneity, this way also contributing to the development of the method. Part IV concludes the thesis. The conclusion is mostly forward-looking: on top of summarizing the thesis' main results and findings, I discuss emerging trends and directions in the study of the economics of preference heterogeneity, and consider how the EP approach may be applied further.

1.4.1. Part I: An Overview and Evaluation of Approaches to Measuring Preference Heterogeneity

Moving on to a more detailed outline of the thesis, the first part consists of two chapters (2 and 3). Both these chapters highlight important features of the three approaches to measuring preference heterogeneity. Chapter 2 is most general, reviewing all three approaches. Chapter 3 scrutinizes stated preferences. After these two chapters we will have a clear picture of the

usefulness (and lack thereof) of two of the three approaches to measuring preferences and assessing preference heterogeneity.

The review in Chapter 2 serves three purposes. Firstly, it brings together the different approaches applied in the economics of preference heterogeneity literature and categorizes them. I give a theoretical background of each method of measuring preferences and assessing preference heterogeneity and illustrate them with a brief review of empirical applications. Secondly, I situate the approaches relative to each other to uncover commonalities between them. The third and main goal of Chapter 2 is to evaluate the three approaches, specifically their underlying measurement methodologies and actual measures of preference heterogeneity. The core criteria are psychometric quality, the measures' reliability (or precision) and validity (or accuracy), and the measures' usability, the potential for applying them in the study of the economics of preference heterogeneity. The RP approach appears rigorous (high validity and reliability) but this comes at the expense of limited usability. Particularly, the RP measures of preference heterogeneity are not widely available and seem highly appropriate for prediction but to have a problem when it comes to distinguishing between the explanans (the thing that explains) and the explanandum (the thing that requires an explanation). The SP approach does better in terms of usability but SP measures may not be particularly valid. Recent work suggests that stated preferences reflect respondents' circumstances rather than their values or preferences. This would make SP measures unsuitable for the study of preference heterogeneity as it becomes impossible to interpret systematic differences in preferences. Finally, the EP approach appears to have some potential but its qualities are uncertain and in need of further demonstration. In particular, almost nothing is known about the psychometric quality of EP measures of preference heterogeneity. I conclude that (as is) there is not one best approach to measuring preference heterogeneity. The approaches face trade-offs between psychometric quality and usability, and their usefulness strongly depends on the specific scientific use to which they are put. Overall, the findings of Chapter 2 lead me to disqualify the RP approach as it has limited usability in the study of the economics of preference heterogeneity, whilst the severity of the measurement problem of the SP approach requires further scrutiny.

In view of this preliminary conclusion, Chapter 3 expands on the evaluation of the SP approach in Chapter 2. It elaborates the critique that SP data are so sensitive to respondent circumstances that they mistake marginal preferences for deep-rooted (cultural) values and attitudinal traits, and finds this measurement problem may account for several instances where individuals profess to have a strong preference for a certain state of affairs or good but

consume only little of this good. I go on to empirically demonstrate the "marginal preferences problem" by applying it to one such case, namely the finding that Muslims and Islamic countries like democracy but have little of it (Bratton 2003; Hoffman 2004; Inglehart and Norris 2003; Jamal and Tessler 2008; Mogahed 2006; Rose 2002; Rowley and Smith 2009). The hypothesis is that this specific paradox may be an outgrowth of SP measures reflecting respondents' circumstances rather than their values or attitudes. If SP measures are prone to elicit marginal preferences—the preference for increasing satiation of an objective given current levels of satiation, rather than deep-rooted attitudinal traits—the preference for satiating the objective in general, the democracy deficit in Muslim-majority countries and Muslims' positive attitude towards democracy are straightforwardly reconcilable and do not constitute a puzzle. Empirical evidence showing that individuals living in undemocratic societies have much more favorable inclinations towards democracy supports this argument and the proposed interpretation of stated values measures as reflecting marginal preferences rather than attitudes. Implication is that SP measures of preference heterogeneity may be invalid, systematically misrepresenting group differences in preferences, and should be used in the study of the economics of preference heterogeneity with care.

Taken together, the first part of the thesis concludes, partly by elimination, that the EP approach is a promising way to study the economics of preference heterogeneity. The method's potential needs to be demonstrated further before final judgment can be passed, however, and I take up this issue in Part II of the thesis.

1.4.2. Part II: Subjective Well-Being and (Heterogeneous) Experienced Preferences

As is, there is some uncertainty or, rather, a basic lack of knowledge about the EP approach and, specifically, its usefulness for the study of the economics of preference heterogeneity. To ascertain how much potential the EP approach really has, this uncertainty needs to be reduced. The lack of knowledge concerns two chief aspects of the EP approach. First, and most directly related to the study of preference heterogeneity, not much is known about the psychometric quality of EP measures of preference heterogeneity. Second, the whole idea of subjective well-being (SWB) is relatively new to economics.

The goal of this part of the thesis is to give more background to the empirical SWB construct and its application in the measurement of preferences and to establish more clearly the EP approach's usefulness for the study of the economics of preference heterogeneity. It consists of two chapters (4 and 5). Chapter 4 deals with questions that are basic to the EP

approach and its chief empirical input, namely what is subjective well-being and how do we measure it. The chapter also expands on the uses SWB can have on top of the measurement of (heterogeneous) preferences. This chapter has a general approach, although for our discussion of applications of SWB the focus is on public policy uses (which includes the measurement of preferences). I discuss the definition of SWB, give several examples of SWB indicators, and discern several interesting applications of SWB in shaping and appraising policy and government action.

Chapter 5 presents a psychometric assessment of both SWB and EP measures of preference heterogeneity. The reliability and validity of SWB indicators is a prerequisite to the use of SWB in the measurement of preferences and heterogeneity therein. I also find, however, that the reliability and validity of the SWB input is not enough to guarantee confidence in the reliability and validity of EP measures of preference heterogeneity. The first part of the chapter therefore surveys the large literature assessing the reliability and validity of SWB indicators, whilst the second part empirically analyses the psychometric quality of preference heterogeneity measured using the EP approach, of which only very little is known. I find that measures of SWB are reasonably reliable and valid—enough so that they can have valuable application in a range of possible uses, not least of which is public policy. To assess the psychometric quality of EP measures I examine the extent to which measured heterogeneity in experienced preferences is consistent across different indicators of SWB, specifically self-reported happiness, an emotional evaluation, and satisfaction, a cognitive evaluation. I find consistency can be quite high with correlations between different EP measures of preference heterogeneity reaching up to 0.95. This bodes well for the reliability and validity of these measures. Psychometric quality appears strongly improved by increasing the number of observations, and this is an important caveat for application of the EP approach to assessing preference heterogeneity.

The main conclusion of this part of the thesis comes in the form of an updated evaluation of the usefulness of the EP approach for the study of the economics of preference heterogeneity.

1.4.3. Part III: Assessments of Heterogeneity in Experienced Preferences

Compared to the first two parts, the third part of the thesis is more concerned with actually studying preference heterogeneity. Chapters 6 and 7 apply the EP approach where the focus is

on heterogeneity in work-related preferences. Preferences of this type have received much attention in economics, but even more in other disciplines.

I start with an assessment of Max Weber's classic thesis concerning a specific Protestant work ethic (Weber 1904/1905 [1930]) in Chapter 6. There are powerful motives for doing so. Firstly, and most importantly, Weber pioneered the study of preference heterogeneity so much so that during the century that has passed since he formulated his thesis, work-related preferences have been at the center of attention in almost anything closely resembling the study of group differences in preferences. Notable examples are McClelland's (1961) work on achievement and also Hofstede's (1980, 2001) framework of five cultural dimensions mentioned above. Secondly, during the last decade, and particularly after 9-11, scientific interest in religion has been revived, and differences in religious background are increasingly recognized as a source of differences in preferences and socioeconomic outcomes (cf. de Jong 2009). Interestingly, the large body of empirical work on work ethic—nearly always using SP methods—has not resulted in any clear body of knowledge on Weber's original thesis on the relation between religion and work ethic. Chapter 6 applies the EP approach to provide novel and systematic evidence on this long-standing issue. I operationalize work ethic intuitively as the psychic cost of joblessness as reflected in unemployment's adverse effect on individuals' experienced utility. Our analysis is based on a large and diverse sample covering almost 150,000 individuals from 82 societies. Empirical results strongly support a Protestant work ethic: unemployment hurts Protestants more, and the psychic cost of unemployment is greater in Protestant societies. Further analysis shows these results are robust and that the effects are independent of social stigma effects as with a moderating effect of peer unemployment (Clark 2003; Clark and Oswald 1994; Shields and Wheatley Price 2005; Stutzer and Lalive 2004).

Chapter 7 deals with differences in value preferences between East and West Germany. Again there are strong motives for studying this subject. The separation and reunification of East and West Germany has provided social scientists with a unique natural experiment (cf. Alesina and Fuchs-Schündeln 2007; Diamond and Robinson 2010: 121; Fuchs-Schündeln 2008; Fuchs-Schündeln and Schündeln 2005; Redding and Sturm 2008; Weil 2009, Chapter 12). Particularly, it gives social scientists the means to study the effect of preferences and informal institutions on socioeconomic outcomes and developments in a setting in which other factors are controlled for. In the case of East and West Germany the outstanding opportunities offered by the separation and the reunification of the former states are enhanced by the availability of large amounts of data, which contrasts with the case of

North and South Korea for example. A further factor driving the importance of East and West Germany as a case study is the fact that decades after reunification substantial economic disparities between East and West Germany remain. With formal institutions being equalized, a typical explanation is that this is due to differences in informal institutions, specifically in values or attitudes. The reasoning is that over 40 years of Communist reign must have instilled in East German minds an economically debilitating cultural legacy. This is often referred to as the "Mauer im Kopf" or "wall in the head." However, empirical work using values surveys to investigate value differences between East and West has so far produced decidedly mixed results (see, for example, Alesina and Fuchs-Schündeln 2007; Corneo and Grüner 2002; Shiller et al. 1991, 1992). I analyze a panel consisting of more than 21,000 individuals and roughly 120,000 observations using multilevel modeling and examine how and to what extent East Germans and West Germans entertain different preferences. Empirical results show that preferences indeed vary between East and West Germans. However, this variation is not in line with the differences associated with the gap in economic performance; if anything, Easterners entertain values more conducive to economic performance. The chapter therefore concludes that the idea of East Germany suffering a Communist cultural legacy appears largely a myth.

1.4.4. Part IV: Conclusion

Part IV of the thesis consisting of one chapter (Chapter 8) concludes. This concluding part gathers together the main results and findings of the thesis and looks ahead at developments in the study of the economics of heterogeneous preferences. It first summarizes four key empirical results. These concern both the approaches by which preference heterogeneity is given empirical content—the reliability and the validity of the resulting measures, as well as the substantive group differences in preferences uncovered in this thesis: Protestantism is associated with higher work ethic and East Germans do not entertain values less conducive to economic performance than West Germans do.

I go on to state two challenges faced by preference research, a general one concerning the endogeneity of preferences and one specific to the EP approach. In addition, I provide an outlook on trends in and directions for research on (heterogeneous) preferences (and their economic consequences). Interestingly, several of the avenues being explored now and in the near future are likely to be multidisciplinary, including, notably, the study of the evolutionary and biological foundations of preferences and variation therein. A most promising future

application of the EP method is in the construction of comprehensive quantitative frameworks of cultural differences between countries and other classes along the lines of Hofstede (1980, 2001) or GLOBE but based on EP measures of preference heterogeneity.

Part I: An Overview and Evaluation of Approaches to Measuring Preference Heterogeneity

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quality as well as its quantity.

- Thorndike (1918: 16)

Chapter 2

Measuring Preferences and Assessing Preference Heterogeneity: A Comparative Review of Approaches

2.1. INTRODUCTION

Economists understand and deal with preferences in three different ways, intuitively, formally, and quantitatively. The intuitive understanding of preferences is as tastes, which involves mental entities including people's attitudes and objectives but also the criteria by which they evaluate outcomes or different states of affairs (e.g. Hirschman 1984; Kahneman et al. 1999; Kahneman and Sugden 2005; Rubinstein 2006). In the context of choice, people choose the alternative they like the most, the alternative from which they expect to derive the most utility. The intuitive understanding of preferences is not dependent on actual choice, however, as when a person likes Picasso's earlier work over his later work, has an attitude towards past events, or with one's favorite color. Choice is decidedly more elemental in the formal understanding of preferences. The formal view sees preferences as a simple mathematical relation: preferences concerning X, a set of commodities, outcomes or courses of action, are given by the relation satisfying the conditions that for any $x, y \in X$, $x \succeq y$ or $y \succeq x$ (completeness) and that for any $x, y, z \in X$, if $x \succeq y$ and $y \succeq z$, then $x \succeq z$ (transitivity). Finally, when economists deal with preferences in a quantitative manner, it tends to be because they want to put a price tag on something or because they want to measure the public's willingness to pay for a particular good or service. Researchers, but also policy makers or marketing agencies, commonly measure preferences to answer practical questions like what is the welfare effect of a reduction in environmental pollution or to predict likely sales of a new product (cf. Payne et al. 1999).

This chapter (and the thesis as a whole) elaborates the measurement of preferences, specifically assessments of preference heterogeneity, i.e. systematic differences in preferences across groups or classes of people. The conceptual backdrop to this quantitative interest in preferences is the view of preferences as tastes. This intuitive understanding of preferences ensures that they are relevant either as guides to human action, or because they matter for individuals' welfare or well-being. Throughout this thesis we use the terms welfare and well-being interchangeably.

As mentioned in the previous chapter, three broad approaches to measuring preferences can be discerned, revealed, stated and experienced preferences (e.g. Fleurbaey 2009: 1054; Frey et al. 2007, 2010; Stutzer and Frey 2010: 23). This chapter categorizes these approaches and takes stock of the way they can and have been applied in assessments of preference heterogeneity. The approaches' defining difference concerns their measurement

methodology, the way in which they give empirical content to preferences and, for our purpose, heterogeneity therein. The first, revealed preference (RP) approach is undoubtedly most familiar to economists, although perhaps not as a tool for assessing preference heterogeneity. Our overview discusses how the traditional RP method has been adapted for application to the study of group differences in preferences. Economists are also relatively familiar with the second, stated preference (SP) approach. As a method for assessing preference heterogeneity it has a long history, though more so in (cross-cultural) psychology than in economics. The third, experienced preference (EP) approach has been developed only fairly recently and is not yet widely known. We therefore discuss it at length.

Next to these three established approaches to measuring preferences and assessing preference heterogeneity, developments in the field of neuroeconomics (Bernheim 2009; Camerer 2007; Camerer et al. 2004, 2005; Glimcher and Rustichini 2004; Rustichini 2005) or genoeconomics (Benjamin et al. 2008; Navarro 2009) are such that, in due course, we may find a fourth approach to measuring preference heterogeneity. Some of the earliest studies of this kind have experimented with administering oxytocin, a neurotransmitter, to subjects to see how it affects their level of trust (as measured using trust games) (see Fehr 2009 for an overview). Directly concerning individual differences in preferences, Sapienza et al. (2009) found that individuals with higher testosterone levels took more risk in a lottery experiment. This suggests that testosterone levels can act as indicators of different preferences for risk. More generally, hormone levels have been linked to personality traits opening up the possibility of biological proxies for personality (Sellers et al. 2007; Newman and Josephs 2009). Finally, recent research has related cultural differences between societies to the likelihood that individuals from these societies carry particular varieties of the serotonin transporter functional polymorphism (5-HTTLPR) (Chiao and Blizinsky 2010; Way and Liebermann 2010). As is, dealing with the neurophysiology of preference heterogeneity is beyond the scope of this thesis. However, we return to this type of work in Part IV of the thesis (Chapter 8) where we give an outlook on trends in and opportunities for research on the economics of preference heterogeneity.

Next to elaborating the theoretical background for the three established approaches to measuring preferences and assessing preference heterogeneity, the overview in this chapter discusses some illustrative empirical applications. In addition, we systematically compare key features of the approaches. This comparison has a theoretical and a practical component. The theoretical component looks at commonalities between the three approaches, specifically their underlying measurement methodologies. Situating the approaches relative to each other, we

find some shared cross-disciplinary elements in all of them. The second, practical component of the comparison aims to evaluate the approaches in relative terms where the overarching concern is their usefulness in the study of the economics of preference heterogeneity. This part of the chapter is most important for the remainder of the thesis.

We formulate two broad criteria by which to evaluate the approaches and their actual measures of preference heterogeneity. At the basis, preference heterogeneity should be given empirical content in a way that renders measures that are meaningful, able to convey precisely and accurately whatever empirical regularities preferences may have at the group level. This is our first criterion. It encompasses the classic psychometric properties of reliability (precision) and validity (accuracy) (e.g. Cronbach and Meehl 1955; American Educational Research Association, the American Psychological Association and the U.S. National Council on Measurement in Education 1999). These two psychometric properties will receive a lot of attention in our discussion, the logic being that if we want to be serious about the measurement of preferences, we should also be willing to assess their quality by these widely accepted standards. Reliability and validity together determine whether a measure is meaningful (or, in the worst case, not worthy of any attention). The idea that empirical measures should be meaningful is internal to the measures, concerning characteristics they themselves may or may not have (though, of course, in relation to the latent construct they are supposed to measure). We refer to this criterion, in short, as the psychometric quality of the measures of preference heterogeneity.

The second criterion on the other hand is external as it relates mostly to the uses to which measures of preference heterogeneity may be put. We label this criterion "usability." It includes, amongst others, a pragmatic assessment of data availability, e.g. for how many groups comparable data on their preferences are readily available or the ease with which such data can be collected. Psychometric quality tends to be a prerequisite for usability. Measures that are too noisy or lack precision offer only limited insight on whichever issue one seeks to examine, although trade-offs exist.

The evaluation finds that the three approaches do not perform equally well on these two criteria. The RP approach appears to have the highest psychometric quality but this comes at the expense of limited usability. Most notably the RP measures of preference heterogeneity are not widely available and insights from the philosophy of science suggest they are appropriate for description and prediction but not for explanation. Some recent work suggests that the SP approach suffers a major problem with the accuracy and precision of its measures. If borne out, this problem disqualifies the use of the SP method in the study of the economics

of preference heterogeneity. Finally, the EP approach seems promising but too little explored. We know next to nothing about the psychometric quality of EP measures of preference heterogeneity, for example.

By necessity, the overall conclusion following this evaluation is somewhat preliminary. In addition, a final evaluation of each of the three approaches requires more information. Exactly how severe is the measurement problem that has recently been uncovered in SP research? And EP measures of preference heterogeneity may have certain advantages over other types of measures, but are they valid and reliable? Chapter 3 and the second part of this thesis (Chapters 4 and 5) are concerned with exactly these questions. Meanwhile, this chapter concludes that there is unlikely to be one best approach to measuring preferences and preference heterogeneity, either theoretically or practically. Analogous to possible trade-offs between psychometric quality and usability, different approaches render measures that are highly suited for one research application but not for another. Implicit in this are practical considerations, not least of which is the cost of measurement.

The organization of this chapter follows the outline given above. Section 2.2 presents our overview, which includes illustrative empirical applications. Section 2.3 situates the three approaches relative to each other, emphasizing their (cross-disciplinary) commonalities. We evaluate the approaches based on our two broad criteria in Section 2.4. Section 2.5 brings together the results of this evaluation, scoring the RP, SP and EP approach in comparative terms. Section 2.6 gives a short summary and conclusion.

2.2. MEASURING PREFERENCES AND ASSESSING PREFERENCE HETEROGENEITY: AN OVERVIEW OF APPROACHES

This section considers the RP, SP and EP approaches in isolation. We discuss some of the underlying theory and the actual measurement methodology involved. As the proof of the pudding is in the eating, we illustrate the three approaches by briefly discussing some empirical applications of each of them.[10]

[10] Related work by Fischhoff (1991) reviews value elicitation. He discerns three paradigms in values research based on their assumption concerning the extent to which it is possible to elicit complete and coherent preferences (see, also, Payne et al. 1999). Our overview and categorization has a different focus, taking work in economics that has taken various approaches to assessing preference heterogeneity as our starting point. This means that our discussion includes three approaches and three different measures of preferences, not just

2.2.1. Revealed Preferences

Theory and Measurement Methodology

The idea of revealed preference is one of the most influential ones in economics (Varian 2006: 99). Originally developed by Samuelson (1938a, 1938b, 1948), RP theory finds that under certain (natural) conditions, individuals' choices do, in fact, coincide with their demand functions. The idea is that economists are not able to observe preferences directly but that on the foundations of RP theory it is possible to connect observed choices to the maximization of an underlying utility function.

An important empirical application of RP theory is to predict behavior. The theory enables researchers to predict people's choices in a given setting from choice behavior originally observed under different circumstances, notably under a different budget constraint. McFadden (1974), for example, shows how data on observed choices can be combined with certain assumptions to estimate probabilistic choice models. Importantly, such estimation does not require knowing or understanding people's motives (cf. Akerlof 2007).

Historically, RP analysis has not been much applied to study systematic variation in preferences. Rather the other way around, RP analysis typically imposes a homogeneity assumption (Manski 2004: 1329). Because most of the time a single individual is not observed under varying circumstances, RP theory can only be turned into a practical tool for predicting behavior by deeming preferences reasonably similar and pooling observations for multiple persons (ibidem). This changed with the advent of behavioral and experimental economics, however.

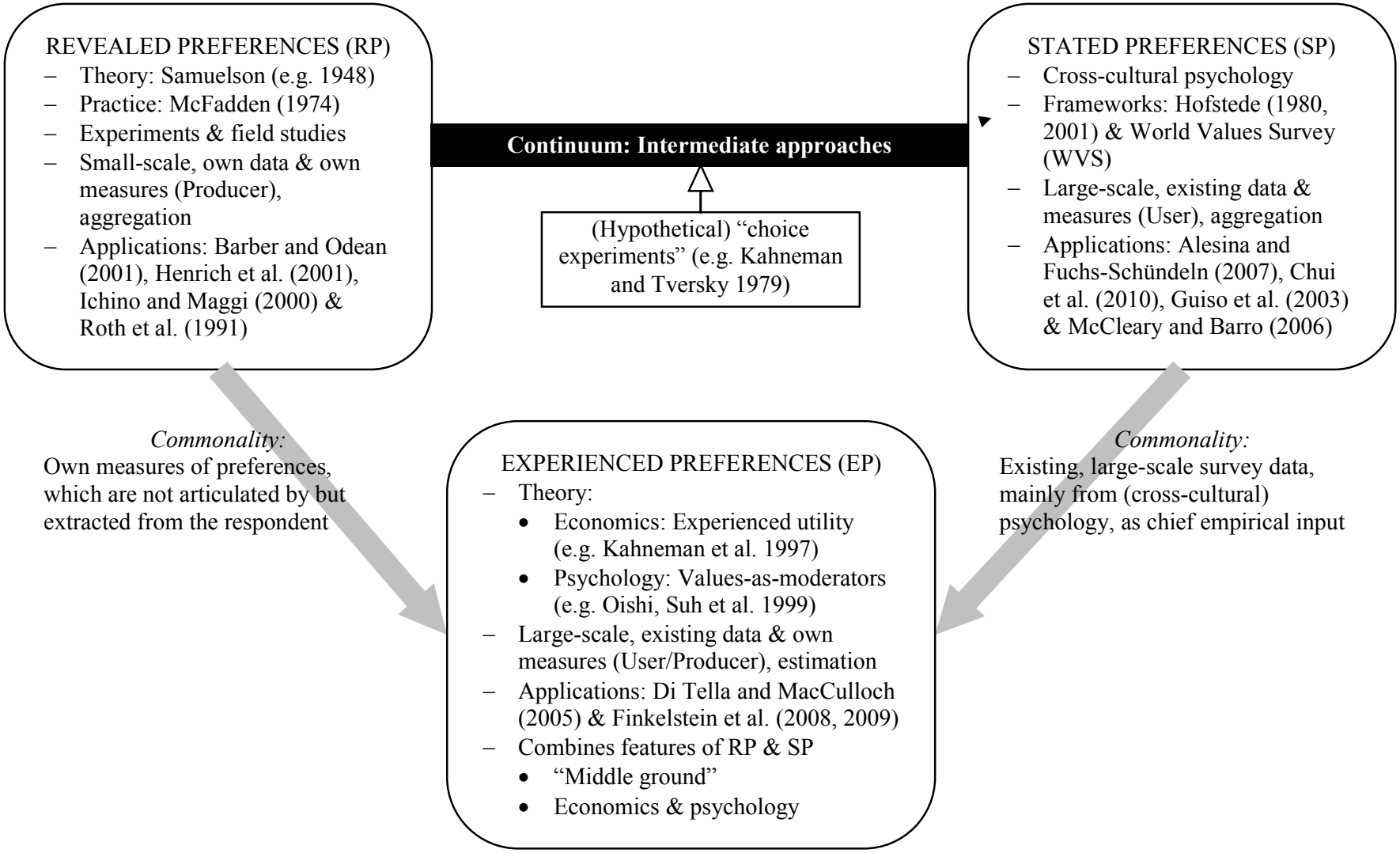
Early work by behavioral economists used choice experiments and economic games—often dictator, ultimatum or public goods games—to study the shape and content of the utility function by means of choices observed in the laboratory (e.g. Camerer and Loewenstein 2003; Levitt and List 2007; Rabin 1998).[11]

---

measures of revealed and stated preferences, but also the recently developed measures of experienced preferences. This way, our overview extends and updates Fischhoff's (1991) survey.

[11] Experimental economics, in contrast, focuses more on outcomes in laboratory markets (cf. Loewenstein 1999; Smith 1989). Because they do not involve random assignment of subjects to a certain treatment, many economic experiments, not least those assessing (heterogeneous) preferences, are actually more correctly referred to as quasi-experiments (e.g. Burnham and Kurzban 2005; see, also, Campbell and Stanley 1966, and van de Vijver and Leung 1997) but we follow tradition here. The advent of behavioral economics also features in the development and application of the EP approach; see the discussion below.

Figure 2.1, Assessing Preference Heterogeneity: An Overview.



To assess systematic differences in preferences between classes (identified beforehand) the researcher simply compares experimental results for a matched sample from each group. If the samples do not diverge on the most important economic parameters, any difference in choice behavior observed for the groups demonstrates preference heterogeneity (cf. Camerer and Fehr 2004; Andersen, Harrison et al. 2010 provide a well worked out example; see also McFadden 1974). In practice, researchers mostly aggregate experimental outcomes and compare means across groups to examine the extent to which preferences differ between them. Applying this methodology, economic experiments become a measurement tool (Croson and Gächter 2010; Guala 2008; Plott and Smith 2008: xi). Somewhat more elaborately, data on experimentally elicited behavioral responses can also be used to estimate a general utility function but with specific parameters that vary across groups (Andersen, Harrison et al. 2008a; Bellemare et al. 2008; von Gaudecker et al. 2010). This compares with the EP approach, which also involves estimating heterogeneous utility functions (see below).

Following the laboratory study of heterogeneity in revealed preferences, some researchers have also found clever ways to study this heterogeneity using field data (cf. Stigler and Becker 1977). The challenge is to have exhaustive controls available. Only when all other differences—think relative prices and budget constraints—have been controlled for, is it possible to conclude that whatever group differences remain must be due to differences in preferences between the groups being compared. The discussion of some illustrative empirical applications to which we turn next includes examples of such work.

Illustrative Empirical Applications

Most RP studies of systematic differences in preferences take the laboratory approach and use economic experiments to measure preference heterogeneity. Differences in proposer offers in the dictator or the ultimatum game, for example, reflect differences in other-regarding preferences and inequity aversion. Games like these, and other economic experiments, have mainly been used to study gender and cross-cultural differences in preferences.

Choices in these games show that men and women are roughly equally altruistic but that offers by women, on average, are more variable and more sensitive to differences in the design of the experiment and to social cues (see Camerer 2003, Croson and Gneezy 2009, and Eckel 2008 for overviews). Cross-cultural economic experiments report more variation (Camerer 2003 and Boyd 2008 survey the literature). Roth et al. (1991), for instance, compare average ultimatum game offers in Israel, Japan, the United States and former Yugoslavia.

They report substantial differences in the amounts offered, but not in rejection rates, which were unaffected by the height of the proposed splits. It therefore seems inhabitants of the four countries entertain different tastes for fairness. Henrich et al. (2001) provide further evidence on cross-cultural differences in inequity aversion. Results from ultimatum game experiments in 15 small-scale societies show offers ranging from 27% to 58%. They also find marked differences in rejection rates, from 0% (out of 51 offers) to 40% (including rejection of more than equal splits).

Researchers also assess gender and cultural heterogeneity in preferences outside of laboratory and game settings in field studies. Gender differences in preferences seem much more pronounced when assessed on the basis of real-life outcomes. Controlling for other factors, women seem, especially, less willing to take risks than men as reflected, for example, in their asset portfolios (Sundén and Surette 1998). In addition, they are less prone to trade their stocks inefficiently often (Barber and Odean 2001). Atkinson et al. (2003) on the other hand find that male and female fund managers do not differ significantly in key characteristics of their funds such as performance and risk.

Studying cross-cultural variation in revealed preferences outside the laboratory is a bit more complicated. The challenge is to distinguish observed economic differences from evidence on preference heterogeneity—the critical appraisal of the RP approach in Section 2.4 elaborates this difficulty. Some studies are able to show that preferences likely vary along cultural lines by adding region- or country-of-origin information to an otherwise standard comparison of outcomes averaged over groups. Such information helps identify variation that has a cultural origin whilst the shared environment controls for other potential influences such as formal institutions. Ichino and Maggi (2000) use this approach, which Fernández (2007) labels the "epidemiological approach," to account for differences in shirking at work in a large Italian bank. Controlling for the location of firm branches and a range of other factors, they find that misconduct and absenteeism are less prevalent among workers born in the North than in the South of Italy. Fernández and Fogli (2006) compare immigrant- and home-country groups in the U.S., finding that the total fertility rate in a woman's country of ancestry predicts differences in average fertility between female immigrants in the U.S.

2.2.2. Stated Preferences

Theory and Measurement Methodology

The SP approach studies the shape and content of the utility function, and possible group variation therein, by simply asking respondents about their preferences. The most common usage of such survey-based data is in valuation exercises when no information on market demand is available. Specifically, the contingent valuation method uses stated demand information (i.e. willingness to pay or willingness to accept) and preferences over consumption bundles to put a price tag on goods, often environmental (dis)amenities, for which a market does not exist (see Arrow et al. 1993, Diamond and Hausman 1994, and Hanemann 1994 for introductions; Cameron 2008 provides an overview). Other work, usually in marketing, uses SP data to predict demand for consumption bundles, for example new goods or services, that differ so much from existing bundles that no relevant RP data are available.

Compared to the RP approach, the SP approach has a less elaborate theoretical framework underlying its application in economic analyses. The main idea is simply that preferences may not be observable to the outside observer but that people know their own preferences and are able articulate these preferences when prompted to do so. As Fischhoff (1991: 835) puts it, "[s]urvey researchers expect their "participants" to provide meaningful answers to items on any topic intriguing them (or their clients), assuming that the questions have been put into good English." Fischhoff (1991: 835) contrasts this with the RP approach and economists expecting ""actors" to pursue their own best interests, thereby making choices that reveal their values, in whatever decisions the marketplace poses (and economists choose to study)." This latter assertion may be more or less true (cf. Fischhoff 1991), as we discuss somewhat further in Section 2.3.

Illustrative Empirical Applications

Key application of the SP approach to assessing preference heterogeneity is in the comprehensive frameworks by Hofstede (1980, 2001), the GLOBE (Global Leadership and Organizational Behavior Effectiveness) project (House et al. 2004) and other such studies that were mentioned in Chapter 1. These large-scale measurement efforts are part of a long tradition of researchers seeking to measure (national) differences in culture, defined as "the collective programming of the mind that distinguishes the members of one group or category of people from another" (Hofstede 2001: 9), and/or in values: "broad tendencies to prefer

certain states of affairs over others" (Hofstede 2001: 5). Values are actually a core element in culture (Hofstede 2001: 10), which for economists reflects "systematic variation in beliefs and preferences across time, space, or social groups" (Fernández 2007: 305). Guiso et al. (2006: 37) similarly take culture to affect values, and values to influence (economic and political) preferences. Attitudes are also common constructs in this literature, and seen as more specific than values, which are more general (Hofstede 2001: 5).

Applications of the SP approach in the study of preference heterogeneity tend to take their input from existing frameworks and surveys rather then taking up measurement themselves; they *use* existing data and measures much more so than that they *produce* new measures using surveys and questionnaire items of their own design (see Figure 2.1). This sets them apart from most RP studies (see above), and also from EP studies, which we discuss below.

Hofstede's (1980, 2001) framework is very popular though mostly outside economics strictly, for instance in international business (e.g. Franke et al. 1991; Husted 1999; Kogut and Singh 1988; Park and Ungson 1997; see Kirkman et al. 2006 for a survey). Recent examples in economics are Beckmann et al. (2008) and Chui et al. (2010). They use Hofstede's quantitative indices of cultural values to account for country differences in asset management and stock market trading strategies. Individualism particularly, by which is meant the extent to which individuals are integrated into groups, affects trading volume, volatility, and safety margins.

The GLOBE framework is increasingly used as a supplement to Hofstede's culture indices. Part of its attractiveness seems to be that it extends the number of cultural dimensions to nine. In addition, it contains both values and practices measures—for a total of 18 country indicators—and covers more societies, 62 in all.

Of the large-scale surveys and frameworks, the World Values Survey (WVS) seems most popular in economics. The WVS uses an extensive questionnaire to probe respondents, inter alia, on their values and attitudes. The complete dataset, including responses for some 350,000 individuals from almost 100 societies, is publicly available (European Values Study Group and World Values Survey Association 2006; World Values Survey Association 2009). These responses form the basis for many studies of the economics of preference heterogeneity. Guiso et al. (2003), for example, use SP data from the WVS to assess religious variation in values and attitudes. They find religion fosters attitudinal traits and value dispositions conducive to economic performance, i.e. higher per-capita GDP and growth. Similarly, McCleary and Barro (2006) report that part of the variation in the value attached to

thrift and hard work can be attributed to religion. These studies reflect an important advantage of the WVS over frameworks like that of Hofstede, namely its flexibility. Preference heterogeneity can be assessed for different groups, such as people with shared religious backgrounds, and not just for different societies (see also Section 2.4 on the usability of the SP approach).

Beyond the WVS, other studies of individual-level SP data have looked at Communism's effect on preferences and at group differences in the preference for redistribution. An example of the former is the study by Shiller et al. (1991), which surveyed random samples of the Moscow and New York populations. Much to their surprise, they did not find much difference between the two groups. Furthermore, the dissimilarities they did find often had a rationale in differing situational factors (cf. Shiller et al. 1992: 179). Corneo and Grüner (2002) confirm this result finding that rational self-interest often accounts for East-West differences in preferences. The study by Alesina and Fuchs-Schündeln (2007), on the other hand, finds support for the claim that Communism has had an effect on values although the effect seems to be waning We discuss this literature in more detail in Chapter 7 where we address the long-standing issue of East-West differences in preferences, which SP measures have thus far been unable to resolve, using the EP approach.

Luttmer (2001), Alesina and La Ferrara (2005), and Keely and Tan (2008) are examples of SP studies of the "taste for redistribution". Luttmer's (2001) analysis finds that individuals feel more negatively about redistribution the higher is the share of people in the community receiving welfare assistance, and more positively the higher is the percentage of beneficiaries that belong to one's own racial group. Alesina and La Ferrara (2005) and Keely and Tan (2008) find that the preference for redistribution varies, amongst others, with sex, race (black vs. white), age, and socioeconomic class.

2.2.3. Experienced Preferences

Theory and Measurement Methodology

Following the intuitive understanding of preferences as tastes, measuring preferences is about finding out what people like and how much they like it. The RP method tries to do this by observing people's choices, with the underlying premise that people choose what they like the most, given their constraints (and avoid what they dislike, again given certain constraints). The SP method is less roundabout, simply asking people how much they like or

dislike particular goods or certain states of affairs. Ultimately, what people like, their preferences or tastes, matters for their welfare.

The EP approach is a recent development—it is not included in Fischhoff's (1991) survey of value elicitation. Its starting point is opposite from that of either the RP or the SP approach. Main empirical input is an indicator of individuals' well-being. Working downwards from this indicator it is possible to say something about people's preferences. To do so, involves relating measured well-being to its potential determinants, which can be any number of things, individual factors such as income or employment status, or country-level variables such as inequality or inflation. The estimated coefficients for the determinants—large or small, positive or negative—show how much people like or dislike something in terms of its contribution to their well-being. The structure of well-being in terms of its determinants reflects people's preferences.

In his *Mathematical Psychics*, Francis Ysidro Edgeworth (1881: 101, italics dropped) famously proposed a "hedonimeter:"

> To precise the ideas, let there be granted to the science of pleasure what is granted to the science of energy; to imagine an ideally perfect instrument, a psychophysical machine, continually registering the height of pleasure experienced by an individual, exactly according to the verdict of consciousness, or rather diverging therefrom according to a law of errors. From moment to moment the hedonimeter varies; the delicate index now flickering with the flutter of the passions, now steadied by intellectual activity, low sunk whole hours in the neighbourhood of zero, or momentarily springing up towards infinity.

For millennia, or at least as long as philosophers have been interested in human well-being, actually measuring how well people are doing seemed beyond the realm of possibilities.[12] Over the last couple of decades, things have changed drastically. As we discuss in more detail in Chapter 4, a large body of research in psychology, and increasingly also in economics, empirically studies happiness and other features of well-being. A straightforward definition of happiness or, less colloquially, subjective well-being (SWB), is as "a broad category of phenomena that includes people's emotional responses, domain satisfactions, and global judgments of life satisfaction" (Diener et al. 1999: 277). The rapidly developing happiness

[12] Colander (2007) recounts economics' historical quest to measure utility.

literature finds a large set of variables to be important determinants of SWB at both the individual and the country level (Argyle 1999, Di Tella and MacCulloch 2006, Diener 1984, 1994, Diener and Suh 1999, Diener et al. 1999, Easterlin 2003, Frey and Stutzer 2002, Headey and Wearing 1992, Layard 2005, 2010, Lyubomirsky et al. 2005, and Nettle 2005 provide overviews). SWB is a function of such factors as marital status and health, but also of economic circumstances like inflation, the unemployment rate, and, of course, income. Such results from empirically estimated happiness functions, showing what people care about in terms of happiness and how much, provide an empirical basis to the EP approach.

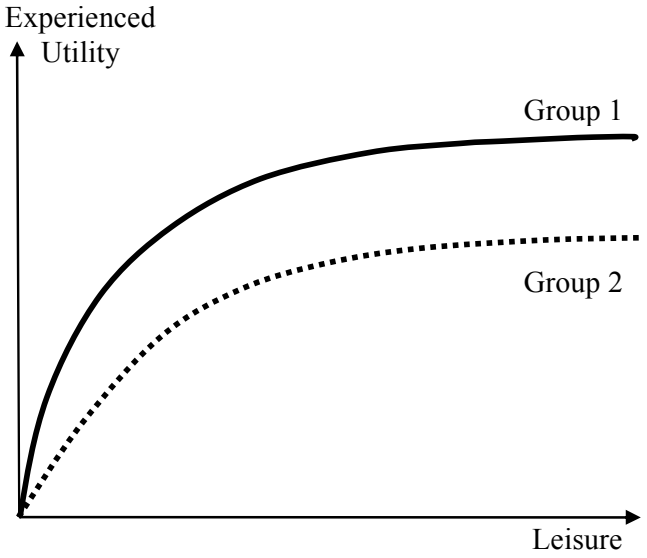
Theoretically, the EP approach is grounded in the work of Daniel Kahneman—psychologist and winner of the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel—and several collaborators on experienced utility (e.g. Dolan and Kahneman 2008; Kahneman 1999, 2000; Kahneman and Sugden 2005; Kahneman and Thaler 1991, 2006; Kahneman et al. 1997). Kahneman is one of the main proponents of behavioral economics (e.g. Heukelom 2009a) and also a member of the Norms and Preferences Network. The EP approach, just like the RP approach and the latter approach's use of behavioral experiments as measurement tools, fits the rise of behavioral economics as a subfield in economics. Kahneman and co-authors distinguish between experienced utility—the "hedonic quality" of an outcome—and decision utility—the "weight of an outcome in a decision" (Kahneman et al. 1997: 375). For Kahneman experienced utility poses a return to the Benthamite conception of utility as pleasure minus pain, whilst decision utility aligns with the traditional utility concept in economics (Kahneman et al. 1997: 375). Irving Fisher's (1918, 1906 [1930]: 43) earlier distinction between desirability or utility and pleasure fits the same mold. Indicators of SWB or happiness can be seen as measures of experienced utility (e.g. Alesina et al. 2004; Di Tella and MacCulloch 2005; Frey and Stutzer 2002; Frey et al. 2009, 2010; Helliwell and Barrington-Leigh 2010a; Rabin 1998; van Hoorn et al. 2010). Using data on self-reported satisfaction or happiness, researchers are able to estimate the shape and content of *experienced* utility functions.[13]

An increasing number of papers uses the EP approach as a valuation tool, for instance, to put a price tag on environmental externalities and other public (dis)amenities that lack a market valuation (e.g. Carroll et al. 2009; Frey et al. 2009, 2010; Luechinger 2009; van Praag and Baarsma 2005; Welsch 2006; Welsch and Kühling 2009). Estimating the SWB effect of

[13] Correspondingly, an alternative name for the EP approach is the "stated happiness" approach (Kahneman and Sugden 2005: 161).

the externality and comparing the size of this effect with that of income allows these valuation studies to calculate the cost of the externality in the absence of market demand information (cf. contingent valuation). The same approach has been used to establish the monetary worth of a person's social capital (Helliwell and Barrington-Leigh 2010b) and to calculate compensatory damages in tort cases (Oswald and Powdthavee 2008). Most of these applications use existing data on SWB and circumstances to produce the preference measures. This way, the EP approach is a *producer* of measures (like most RP studies) and a *user* of data (like most SP studies) (see Figure 2.1) We discuss these applications of the EP approach in valuation exercises such as the ones just mentioned in more detail in Chapter 4, which contains an extended survey of possible public policy uses of indicators of SWB.

Figure 2.2, Experienced Utility Functions with Preference Heterogeneity.



Valuation or measuring preferences is of course not the same as assessing preference heterogeneity. Such heterogeneity is prominent in the psychological SWB literature, however. The values-as-moderators framework of SWB (e.g. Malka and Chatman 2003; Oishi, Lucas et al. 1999; Oishi, Suh et al. 1999; Suh et al. 1998) holds that values moderate the happiness effect of various situational factors. A large body of evidence confirms that the empirical structure of happiness varies systematically with people's self-reported dispositional traits at both the individual and the societal level (e.g. Diener and Diener 1995; Diener et al. 2000; Kuppens et al. 2008; Kwan et al. 1997; Malka and Chatman 2003; Oishi, Lucas et al. 1999; Oishi, Suh et al. 1999; Park and Huebner 2005; Schimmack et al. 2002; Suh et al. 1998; see Diener et al. 2003, Diener and Oishi 2004, and Oishi 2010 for overviews).

Figure 2.2 suggests that, relative to the other approaches, the EP approach looks most directly at the shape (and content) of the utility function. This derives from the fact that the EP approach actually estimates an experienced utility function taking all individual observations belonging to a particular class as its input, whereas the RP and SP method rely on aggregation of choices made or answers given by these same individuals. Implication is that the EP approach requires many individual observations. Furthermore, these observations need to exhibit a reasonable amount of variation in the consumption variable. If all individuals belonging to a particular group consume the same amount of leisure, it becomes impossible to get a general estimate of the preference for leisure for the group. The more variation of consumption within the group, the more completely the shape of the representative utility function for the members of this group can be ascertained. Measuring preferences for aggregate outcomes, say, equality or inflation, requires similar variation in the independent variable. Typically, this involves looking at individuals in different countries for several years, during which the level of equality and inflation has varied within the countries. Section 2.4 discusses the implications of this need for variation for the usability of the EP approach.

Illustrative Empirical Applications

EP analysis does not have a long history, and accordingly its applications have not yet developed into a systematic body of research. As just indicated, through the values-as-moderators framework of SWB (e.g. Oishi, Lucas et al. 1999; Oishi, Suh et al. 1999), psychology has some tradition in assessing preference heterogeneity by way of group-variation in happiness functions. A typical empirical result in this literature is that the structure of happiness in individualistic countries differs from that in countries deemed collectivistic (cf. Hofstede 1980, 2001), for instance with regard to the happiness effect of self-esteem (Diener and Diener 1995) or the happiness bonus married individuals enjoy (Diener et al. 2000).

Some of the group heterogeneity economists have taken up studying concerns the preference for equality and for income. The groups they have focused on include different countries but, particularly Europe versus the U.S., left-wingers versus right-wingers, and people in poor health versus people in good health. Alesina et al. (2004) estimate heterogeneous happiness functions to shed light on differences in the size of the welfare state between Europe and the United States and on partisan preferences. They report that inequality has a larger negative effect on SWB in Europe than in the U.S. and that left-wingers are hurt more by inequality than right-wingers are but only in Europe. Relatedly, Aknin et al. (2010)

document different experienced preferences for prosocial spending (donating to charity) across a sample of 136 countries. Partisan differences are also reported by Di Tella and MacCulloch (2005). They test models of partisan business cycles and find that the happiness of left-wing (right-wing) individuals is lowered more by unemployment (inflation) than by inflation (unemployment). Di Tella et al. (2010) report further left-right differences in the experienced preference for status and income. Lelkes (2006) and Finkelstein et al. (2008) are some other EP studies of heterogeneity in the preference for income. The former reports that the effect of economic variables including income on happiness is smaller among the religious than among the non-religious. Finkelstein et al. (2008) apply EP analysis to the study of the health-state dependence of the utility function (see, also, Finkelstein et al. 2009). They find that the marginal utility of consumption, as measured by the effect of the individual's consumption on his or her happiness rating, declines with deteriorating health. Finally, also concerning income, the exploratory study by Clark et al. (2005) looks at sources of heterogeneity in the transformation of income into financial domain satisfaction in a sample of European countries. They do not identify groups beforehand, however, but apply a latent-class technique. Class membership is based on the way income affects satisfaction with one's finances and is predicted by nationality amongst others.

2.3. COMMONALITIES IN APPROACHES TO MEASURING PREFERENCES AND ASSESSING PREFERENCE HETEROGENEITY

From the survey of the RP, SP and EP approaches and their most important features in Section 2.2, it may seem they have a fundamentally different outlook on measuring preferences. The measurement methodologies that are the foundation of the distinctive approaches also share a lot, however. This section takes stock of the features two or more of the measurement methodologies have in common, and uses the gathered insights to situate the three approaches relative to each other (see Figure 2.1). We identify two broad areas of unity, the influence of psychology, which we discuss in the first part of the section, and the relation between the subject and the observer in the three approaches, which we discuss in the second part. We add a third part that deals exclusively with stated versus revealed preferences as they appear to have some particularly significant commonalities. The latter analysis reveals that, though there are important differences, the grounds for distinguishing between RP and SP

measures can be very fuzzy. The section ends with a brief conclusion. Included in Figure 2.1 are the most important findings of the discussion in this section.

2.3.1. The Influence of Psychology

A first important factor tying the three approaches to assessing preference heterogeneity together is psychology, specifically the approaches' use of theory, techniques, and even data from this field (see Chapter 1). The SP and EP approaches not only take theoretical and empirical/methodological input from psychology, they also originate largely from within psychological research. As mentioned in Chapter 1, psychologists have a long history of studying values and culture, for which they gather data through questionnaires (e.g. Allport and Vernon 1931; Kluckhohn and Strodtbeck 1961; Morris 1956). Much research in (cross-cultural) psychology works with small, purposefully selected samples but the results of some large-scale, nationally representative projects (e.g. Hofstede 1980, 2001) have an obvious appeal to many researchers as well. In a similar fashion, SWB research takes its data from survey items asking individuals how happy they are or how satisfied they are with their lives. Chapter 4 discusses these measures in detail, but note already that values surveys such as the WVS often include one or more questions probing respondents about their happiness or satisfaction. Psychologists have been studying SWB for over five decades and economists have followed up on their work, extending and adapting it to suit their needs, and applying it in economic analyses.

In the RP approach, in contrast, choice takes center stage. This way, RP theory allows researchers to conduct economic analyses without the need to refer to mental states such as people's motives. Notwithstanding, psychological research does affect the RP approach, but its influence derives neither from data nor from theories in psychology. Rather, the influence is limited to methods, concerning only the use of experimental techniques.[14]

As with values and cross-cultural research, psychology has a long-standing tradition in behavioral experiments, even experiments on economic topics such as the law of demand and supply and the principle of diminishing marginal utility (Moscati 2007 provides a historical account). Experimental and behavioral economists have taken psychologists' experimental

[14] At the same time, psychological research may have inspired economists to devote more attention to the content of the utility function in the first place (cf. Camerer 1999; Camerer and Loewenstein 2003; Heukelom 2009a; Mullainathan and Thaler 2000; Rabin 1998, 2002; Sent 2004; Simon 1987).

techniques, again adapting them—notably introducing explicit and often substantial monetary incentives—and using them in their analyses of, for example, the functioning of markets.

2.3.2. The Subject and the Observer

The second important commonality between the RP, SP and EP approaches concerns the relation between the researcher and the people for which group variation in their preferences is assessed. For the measurement of stated preferences, researchers rely on respondents to state their own preferences—this under the assumption that they are able to do so (see above, and the ensuing evaluation of approaches and their measurement methodologies in Section 2.4). In both the RP and the EP measurement methodology, in contrast, preferences are extracted by an outside observer.

RP theory finds that if an agent consistently chooses the option that is (weakly) preferred to all other available options, it is possible to rationalize observed choices as maximization of a utility function. This mapping of choices onto preferences does not demand effort from the subject him- or herself nor does the actual mapping place a cognitive challenge on the subject; the researcher can extract preferences externally. EP measurement similarly extracts preferences but requires even less structure than RP measurement because respondents are only required to report their level of happiness; the relation between situational factors is made in second instance by an external observer (cf. Di Tella and MacCulloch 2006: 40; see, also, Frey et al. 2009, 2010 and Maseland and van Hoorn 2010a).

For both approaches, it is possible to question their key presumption: do individuals indeed always choose the best available alternative; and are people indeed able to judge how happy they are in a way that renders a meaningful measure? Most people actually agree that preferences, whilst rationalizing observed choices, may not reflect individuals' "true" preferences (e.g. Beshears et al. 2008; Dolan and Kahneman 2008; Kahneman and Sugden 2005; Kőszegi and Rabin 2007, 2008). Where the RP method breaks down, the SP and EP method can be used to supplement it (ibidem). In fact, a key consequence of bounded rationality is that decision utility and experienced utility may not coincide. Measures of SWB have encountered even more skepticism, an issue to which we return in the next section and, especially, in Chapter 5. For now, the bottom line is that both the RP and the EP approach measure preferences external to the agents whose preferences are being assessed.

2.3.3. Special Focus: Distinguishing Stated and Revealed Preferences

How Stated and Revealed Preferences are Similar

Though not evident at first sight, there is substantial overlap between the RP and the SP approach, specifically between their respective measurement methodologies. Typical RP analysis of preference heterogeneity involves small samples of individuals from different groups participating in economic experiments such as decisions under risk or bargaining games. In these laboratory experiments individuals choose, for instance, between different (binary) lotteries or how much of a given monetary pie to offer to a third party.

A conceptual step in clarifying the relation between revealed and stated preferences is to view choices made in these experiments as choices between consumption bundles.[15] This is most obvious in the choice between prospects; thus Kahneman and Tversky (1979) ground their prospect theory in evidence from individuals' choices in lottery experiments, asking which option people prefer, for example, win $4,000 with probability 0.8 and $0 with probability 0.2, or win $3,000 for sure (Kahneman and Tversky 1979: 266). The two consumption bundles individuals are asked to compare in this experiment are the certain sum of $3,000 and the ticket to a lottery with an 80% chance of winning $4,000. Since SP analysis is also very much about choices between consumption bundles, it seems difficult to draw a strict dividing line between the RP approach and the SP approach.

We propose seeing the difference between the two approaches as a matter of degree (see Figure 2.1). At the extremes of our suggested continuum things may be clear-cut: a game calling for the division of a hundred dollar bill renders a clear case of a revealed preference, whilst asking people whether they agree or disagree that work not leisure is what makes life worth living, renders an obvious example of a stated preference.[16] In other instances, the distinction is blurry; the hypothetical pairwise choice experiments Kahneman and Tversky (1979) relied upon for their seminal work, for example, may be placed halfway the

[15] This discussion pertains to preferences revealed in laboratory settings and not to natural experiments that seek to uncover preference heterogeneity in a (market) environment on which enough information is available for the researcher to statistically control for other factors possibly affecting the behavior of individuals and the choices they make (e.g. Ichino and Maggi 2000; Fernández and Fogli 2006).

[16] This formulation draws on an actual SP item included in the WVS. The item and example appear again in Chapter 3 and Chapter 6.

continuum.[17] A formal distinction may be whether the choice between consumption bundles is phrased in such a way that it is eligible for actual implementation. The experiments in Kahneman and Tversky (1979) allow for this kind of implementation so that, applying this criterion, they should be interpreted as measuring revealed preferences. A typical SP questionnaire item, on the other hand, offers no means for the researcher to confront the subject with the tangible effects of his or her choice. Consider again a survey item concerned with how much value respondents attach to work relative to spare time. How would we go about making the answer to this question consequential for the respondent? Because the researcher has no control over subjects' actual work and leisure patterns this item renders a stated preference and not a revealed preference.

A second similarity between the two approaches is that both apply aggregation in measuring preference heterogeneity. Individual choices, observed in laboratory experiments or in the setting of a quasi-natural experiment, are aggregated at the level of the group (for example a nation) to which the individual is deemed to belong. Preference heterogeneity is assessed by averaging and comparing outcomes, but the approach retains its basis in individual-level observation. The SP approach applies the same basic procedure, taking averages of individual responses to survey questions and comparing groups' mean scores on these items. As indicated above, this *aggregation* sets these two approaches apart from the EP approach, which uses *estimation* to assess the shape (and content) of the experienced utility function for groups as a whole. The difference between aggregation and estimation surfaces again in Section 2.4, then in the context of the usability of measures of preference heterogeneity.

How Stated and Revealed Preferences Differ

The above discussion of how stated and revealed preferences are similar invokes the question on which point or points the two differ. One theoretical view of the distinction between revealed and stated preferences is as actual choices (RP) versus intentions (SP). Accordingly, the main feature distinguishing revealed preferences from stated preferences appears to be the practical implementation of the measurement methodology. Firstly, economic experiments tend to have explicit, and not seldom substantial, monetary incentives

[17] Smith (1989: 162-165) uses the term choice experiment (or related terminology such as choice survey) in opposition to economic experiments that involve markets. Kahneman-Tversky type, discrete choice experiments are also commonly used to estimate models of demand for non-existing commodities (see Louviere and Hensher 1983 and Louviere and Woodworth 1983 for early examples).

(cf. Loewenstein 1999; Camerer and Fehr 2004; Levitt and List 2007), which surveys of stated preferences do not. Designing laboratory games and behavioral experiments this way clearly conveys to participants what the (monetary) consequences of their choices between consumption bundles are. In surveys, the choice-consequence link is much weaker. At the same time, this distinction still involves some judgment. Particularly, it begs the question how large the explicit incentives need to be before we can talk about revealed preferences and actual choices rather than stated intentions and stated preferences. A dime does not seem enough for U.S. participants, but what about a dollar, or two, or ten? What is more, when respondents do not face explicit incentives does not mean that their choices are without consequences—implicit incentives abound (cf. Loewenstein 1999: F31; Levitt and List 2007).

The above distinction fits with the proposed continuum between revealed and stated preferences. The same can be said about the second theoretical difference, which concerns extensionality (cf. Kahneman et al. 1999; Kahneman and Sugden 2005). Revealed preferences are extensional as they face real-world constraints whilst stated preferences may report only about a particular individual's state of mind. The difference, as Kahneman and Sugden (2005: 164) explain it, is between preferring to climb one mountain to climbing another mountain versus having a positive attitude towards one mountain and a negative attitude towards another.

A third, practical difference is that economists typically conduct experiments with laboratory games such as the trust game repeatedly. This gives subjects the chance to learn. They receive direct feedback from prior experimental outcomes and can use this information in subsequent rounds. In contrast, there does not seem to be much value in asking people whether they agree with a certain statement more than once within a short time span.[18] Furthermore, experiments tend to take place in a laboratory setting allowing more control over the situation (e.g. use of subtle visible cues) than is the case with, for example, a paper-and-pencil questionnaire completed at home or a telephone interview.

The most important difference, however, is that not all survey questions are eligible for incorporation in an experimental design with explicit monetary incentives. Researchers often rely on stated preferences when data on market demand is unavailable and, similarly, mimicking the relevant consumption bundle in an experimental design may be impossible. Both the continuum in Figure 2.1 and the extensional versus non-extensional distinction

[18] Researchers often repeat a question or use two highly similar questions within a single survey to assess the item's reliability. See the next section for more details.

reflect this. The evaluation of approaches and measurement methodologies in the next section pays explicit attention to the range of preferences for which laboratory experiments may offer a suitable measurement tool.

2.3.4. Discussion and Some Integration

The three ways of giving empirical content to preference heterogeneity that can be discerned, RP, SP and EP measurement share some important features. This is not to say that the approaches are all close, or are all equally closely related, but some striking commonalities (and similarities) exist. At a fundamental level, the joint focus on individuals' choices between consumption bundles ties together revealed and stated preferences. In addition, an established body of knowledge in psychology concerning theory, data, and techniques informs all approaches.

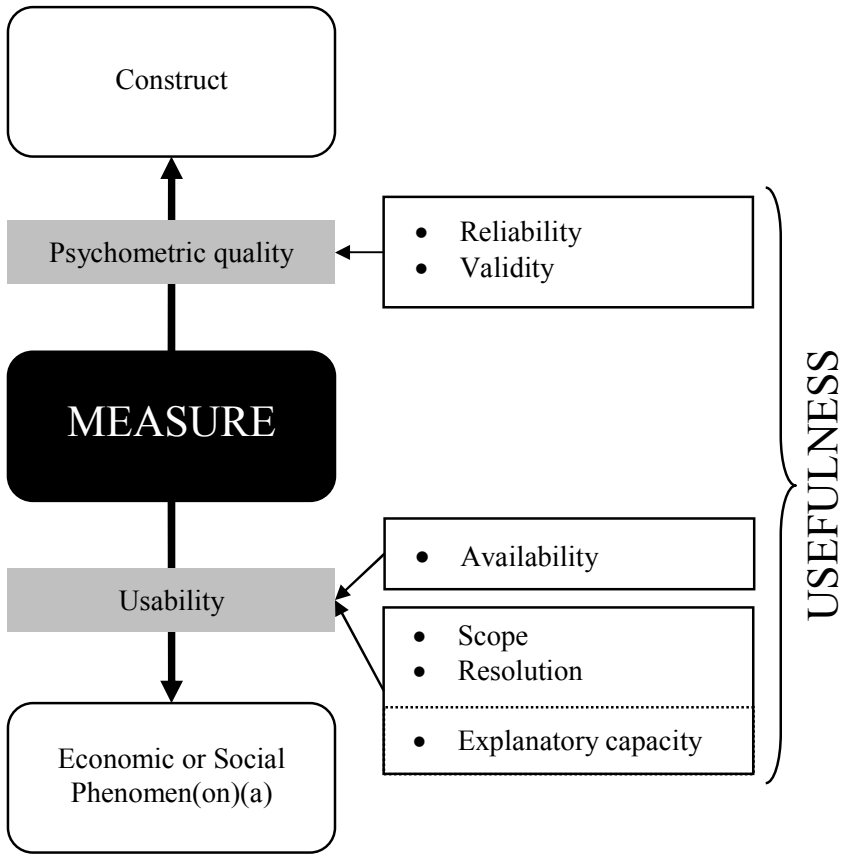
The EP approach emerges as a combination of the RP and the SP approach. Figure 2.1 reflects how the EP approach can be seen as holding the middle ground between these other two approaches, sharing important features with each of them: like the RP measurement methodology, it measures preferences external to the subject, and like (cross-cultural) SP methods, it takes data obtained through large-scale surveys as its empirical input.

2.4. EVALUATING APPROACHES AND MEASURES

How useful are the three types of measures of preference heterogeneity? In this section we critically evaluate the three approaches to measuring and assessing preference heterogeneity, specifically, the measurement methodologies on which they are founded and the actual measures. We apply two core criteria, which we disaggregate in sub-criteria. Whilst this section presents details of the evaluation, the next section brings our findings together to score the three approaches on their usefulness for studying the economics of preference heterogeneity. We start our evaluation with a discussion of the criteria, which we apply later on. Figure 2.3 graphically displays the criteria and sub-criteria that we use in our evaluation, and how they relate to each other. We should note from the start that throughout this thesis we often speak of the RP, SP and EP approaches as having certain properties, i.e. scores on our criteria, whilst formally it are not the approaches that have these qualities but the actual

preference measures (see, for example, Thompson and Vacha-Haase 2000). For our purpose, nothing material is lost by not emphasizing this distinction.

Figure 2.3, A Schematic Overview of the Usefulness of Measures of Preference Heterogeneity.



2.4.1. Criteria for Evaluation

Formal Psychometric Quality: A Short Introduction to Reliability and Validity

A vital prerequisite for the empirical study of preferences is that the quantitative measures of the preferences to be studied are meaningful. This is our first core criterion by which to evaluate the overall usefulness of the three approaches to measuring preferences and preference heterogeneity. From this logic it follows that we should assess the quality of measures of preference and heterogeneity therein by accepted standards and see to what extent they stand up to critical scrutiny. In assessments of the quality of numerical measures in psychology and the social sciences, two key features or psychometric properties stand out.

Below we review some essentials of psychometrics, elaborating these two key properties and drawing analogies with the economics literature.

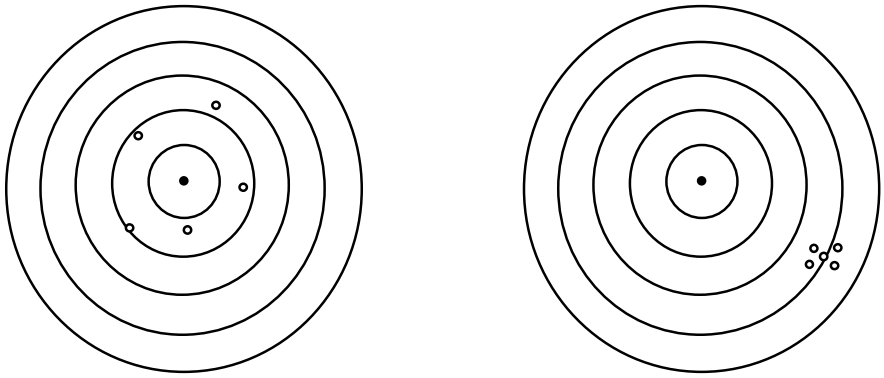
Reliability and Validity as Criteria of Psychometric Quality

The two key properties that assessments of the quality of numerical measures in psychology and the social sciences focus on or are reliability and validity. An indicator's reliability denotes its consistency and its ability to give the same results in repeated measurement. An indicator's (construct or measurement) validity refers to the extent to which the indicator is indeed able to capture the construct it purports to measure. Reliability and validity are alternatively referred to as precision, reproducibility, repeatability or variable error, and accuracy or constant error respectively. Reliability and validity are standard quality requirements for psychometric measures of latent constructs and are extensively dealt with in any textbook on measurement and methods in psychology (see, for example, Hopkins 1998 and Thorndike 2004). The classic references are Cronbach and Meehl (1955) and the handbook on Standards for Educational and Psychological Testing developed by the American Educational Research Association, the American Psychological Association, and the U.S. National Council on Measurement in Education (last revised in 1999). Together, the reliability and validity of a particular indicator reflect the extent to which it captures something meaningful.

Applications of the reliability and validity criteria run widely. Psychological and educational tests, for instance, are extensively evaluated on criteria of reliability and validity. This holds for commonly used tests like those aiming to measure individuals' intelligence (IQ tests) or tests of school children's mathematical and verbal skills, including the American Scholastic Aptitude Test / Scholastic Assessment Test or SAT (now SAT Reasoning Test).

Theoretically, reliability and validity are clearly distinct concepts. The difference is often explained using a target analogy (Figure 2.4). Measurement is more valid or accurate the closer shots are to the actual target center, the bull's-eye. In addition, separate shots may well miss the target, but as a whole they clearly identify the actual target. This is depicted in the left-hand side of Figure 2.4. Reliability or precision, on the other hand, refers to the variation in the various shots. The closer together different shots are, the higher is precision. Even if the shots miss the actual target, if they are grouped closely, precision is high. The right-hand side of Figure 2.4 illustrates this. Reliability is seen as a prerequisite for validity: if individual shots are not close together, they cannot all be close to the bull's-eye either.

Figure 2.4, Accuracy or Validity versus Precision or Reliability in Psychometric Evaluation: The Target Analogy.



Validity or Accuracy *Reliability or Precision*

Reliability and validity can be evaluated in various manners. An oft-used criterion by which to judge reliability is to look at test-retest correlations. This test applies the same instrument for measuring a certain construct to the same set of subjects, but at different occasions. The idea is that reliability requires that at each occasion the same variation in scores is found. Hence, if the correlation between these two measurements is high, we can conclude that the instrument is reliable. A second common reliability criterion concerns instruments' internal consistency. Here the idea is that if an instrument consists of different sub-items, each of which targets the same underlying construct, scores on these sub-items should be strongly correlated. To illustrate, questions from the SAT Reasoning Test could be split evenly to see whether students' test scores vary a lot between the two halves. If they do, the test does not seem internally consistent, whilst a strong correlation between the fifty-fifty splits bodes well for the test's internal consistency and therefore its reliability.[19] To measure internal consistency researchers often use Cronbach's alpha (α). This statistic is calculated from the pairwise correlations between sub-items and can vary between 0 and 1. Values for Cronbach's alpha are interpreted differently. Nunnally (1967: 226) finds that in "the early stages of research on predictor tests or hypothesized measures of a construct, alphas of 0.60 or 0.50 will suffice". Basic and applied research, however, require higher alpha's, in the 0.80 and 0.90-0.95 range (ibidem). Alpha's can also be too high. In these cases, one or more of the sub-items may be redundant.

[19] Here the caveat given above applies: formally, it is not the test that is reliable but the results obtained from applying the test, i.e. the test scores, that are reliable (cf. Thompson and Vacha-Haase 2000).

Tests of validity, not unlike validity itself, come in diverse forms. For practical purposes, validity is often divided in different types with accompanying tests. Two of the most prominent types are convergent validity and discriminant validity. Convergent validity refers to the correspondence between the indicator of interest and other indicators purporting to measure the same construct. If the measure of interest is to have convergent validity the results obtained should converge on the results obtained using these other measures. Discriminant validity considers how well the indicator correlates (or, usually, fails to correlate) with measures that capture something different.

In practice, it is not always possible (nor necessary) to draw a clear distinction between reliability and validity. As a guideline, reliability checks focus more on measurement of the *same construct* using the *same method*. Validity checks on the other hand focus more on measurement of either *different constructs* or using *different methods* or both. Construct validity, for instance, is assessed by looking at the correlation between the original indicator and some other measure of the same construct but obtained using a different method; for discriminant validity the constructs differ and occasionally the method as well though this is not necessary. Even with these guidelines in mind, the distinction between reliability and validity can be difficult to make. Hence, where possible we distinguish between the two quality concepts, but mostly we refer to reliability and validity jointly as the psychometric quality of an indicator.

Reliability and Validity in Economics

Economics pays less explicit attention to the reliability and validity of its empirical data than the other social sciences do, Morgenstern's (1963 [1950]) classic work on the accuracy of economic observations notwithstanding. The concept of validity enters economic discourse mostly in the context of experiments in economics. In this literature, a distinction is made between internal validity and external validity, which are both concepts by which researchers judge the quality of economic experimentation. Internal validity then refers to the ability to draw causal inferences from a particular research and external validity to the ability to generalize from experimental results (cf. Loewenstein 1999; see, also, Leviton 2001 and Mark and Reichardt 2001). Interestingly, the terms internal and external validity were originally introduced by Campbell (1957) where he applied Cronbach and Meehl's (1955) concept of (measurement) validity to psychological experiments (Heukelom 2009b; cf. Leviton 2001; Mark and Reichardt 2001).

More instructive economic equivalents to the psychometric properties of reliability and validity may be found in the criteria economists have applied to judge the quality of theoretical models. Stigler (1950: 392-396) has formulated some of the most influential of these criteria (see also Stigler 1965). He finds that economists chose between theories on the basis of their (i) generality, (ii) manageability, and (iii) congruence with reality. Of these criteria the first two are formal (Stigler 1950: 394) and therefore difficult to relate to psychometric discourse. The third criterion is empirical, however, and refers to theoretical models' ability to "systematize and "explain" a portion of the empirical knowledge of the times" (ibidem). This is not a criterion we can directly apply to quantitative measures and their quality, but it does echo the idea that there should be a strong link to whatever researchers are doing (measuring constructs, theorizing causal relations, et cetera) and observable reality.

The same link can be found in Friedman's (1953) essay on the methodology of positive economics, and the recent update of the Stigler criteria by Gabaix and Laibson (2008). Friedman (1953: 4) posits a simple quality criterion namely that the "performance" of positive economics must be judged "by the precision, scope, and conformity with experience of the predictions it yields." Compared with Stigler (1950), Friedman (1953) emphasizes the empirical part of theories—for him the prediction criterion is the only criterion. Gabaix and Laibson (2008) extend Stigler's (1950) three criteria to seven, partially because of a fine-graining of these original criteria. The criteria Gabaix and Laibson (2008) propose, are: (i) parsimony, (ii) tractability, (iii) conceptual insightfulness, (iv) generalizability, (v) falsifiability, (vi) empirical consistency, and (vii) predictive precision. Of these, only the empirical consistency criterion turns out to fit psychometric discourse, referring simply to the extent to which models are consistent with the data. The empirical consistency criterion thus closely resembles Stigler's (1950) congruence with reality criterion but Gabaix and Laibson (2008) add to this that empirical consistency can be achieved through making predictions that are empirically verified, this way partially incorporating Friedman's (1953) view on what constitutes a good economic theory.

Of course, there is a fundamental difference between criteria applied to models and criteria applied to empirical measures. Nevertheless, the above discussion shows that the ideas of reliability and validity are not alien to economics, even if the terms themselves are perhaps not very common. The quality criteria for economic models emphasize the formal, but the general gist fits the psychometric discourse very well. Reliability and validity, in turn, seem valuable concepts to use whenever economics research involves the measurement of some

(latent) construct. This would seem to hold for familiar economic indicators such as GDP or the unemployment rate as well, even though it may often be unfeasible to actually assess the validity and reliability of these indicators (but see Morgenstern 1950 [1963]).

Usability

Our second core criterion concerns the measures' usability. Every measure or indicator has a degree of reliability and validity. This holds independent of the uses to which the measures are put, the economic and social phenomenon or phenomena the researcher seeks to study with these measures. Usability on the other hand is a property of a measure that tends to be external to the measure, emerging only in relation to its empirical application. Compared with psychometric quality, usability is a much more pragmatic criterion, although the former tends to be a prerequisite for the latter. Overall, usefulness of measures of preference heterogeneity is a compromise. Researchers may accept lower reliability or validity in exchange for higher usability, for example. We subdivide usability in four further criteria: (i) availability, (ii) scope, (iii) resolution, and (iv) explanatory capacity. These sub-criteria similarly involve trade-offs. The first three sub-criteria are related and in the evaluations below we discuss them jointly.

Availability

Availability is the most practical of our usability criteria. It is not a criterion internal to an indicator like reliability and validity, but it is also not strictly dependent on the specific question a researcher seeks to address with a particular measure of preference heterogeneity. We use the term "availability" to signify two things. First, it refers to the extent to which a particular approach—RP, SP, or EP—has comprehensive data available for use in the study of the economics of preference heterogeneity. Second, it denotes the ease with which such data can be obtained using a particular method. There are clear differences between the approaches on this count. As indicated earlier, some types of studies are very much involved with both measuring preferences and assessing heterogeneity therein (RP studies; user and producer), whereas others mainly assess heterogeneity in existing preference data (SP studies; user) (see Figure 2.1).

Scope

The "scope" sub-criterion concerns the extent to which measures of preference heterogeneity are able to reflect comprehensively whatever empirical regularities preferences may have at

the group level. The basis of the evaluation of the approaches on this count therefore is whether measures are able to measure heterogeneity in all preferences of interest to economists (and social scientists at large). Since scope indicates the availability of measures of certain types of preferences it is closely related to the availability sub-criterion. The difference is that the approaches differ in scope on the count of their nature and not for some pragmatic reason such as ease of measurement.

Resolution

In measurement, resolution or resolving power refers to the smallest difference a measuring instrument is able to detect. We use it to denote the smallest observational unit for which heterogeneous preferences can be measured, the level of disaggregation that a particular method for measuring preference heterogeneity allows. Resolution is again a property innate to the approaches although it ultimately boils down to the (non-)availability of disaggregated measures of preference heterogeneity. Similarly, it is the scientific use to which the measures are put that determines the relevancy of the "resolution" sub-criterion.

Explanatory Capacity

Our final, "explanatory capacity" sub-criterion follows directly from the application of measures of preference heterogeneity in the study of economic differences and their sources (e.g. Guiso et al. 2006; Phelps 2006; Phelps and Zoega 2009; Tabellini 2008, 2010; Williamson 2009; and Williamson and Mathers 2010; see Beugelsdijk and Maseland 2011 and de Jong 2009 for overviews). Simple logic dictates that for this purpose the chief requirement for a numerical indicator of preference heterogeneity is that it actually allows the researcher to explain observed economic disparities on the basis of measured preference heterogeneity. A measure's capacity to explain on the other hand is not innate.

The nature of explanation has been widely discussed in the philosophy of science (see, for example, Hempel and Oppenheim 1948 and Hempel 1965; Salmon 1989 gives a historical overview of the debate). A clear requirement for scientific explanation is that it should not involve a logical necessity. Hanson (1958) and Nozick (1981), in particular, warn against circular reasoning in scientific explanation. They find that "what requires explanation cannot itself figure in the explanation" (Hanson 1958: 120) and that "fundamental explanation of [a] property […] will not refer to other things with that very same property; the possession and functioning of that property is what is to be explained" (Nozick 1981: 632). A typical

example is that an explanation of why a given table is purple should not make an argument that this is because the atoms that comprise the table are purple (cf. Hanson 1958: 121).

For empirical indicators, including those capturing preference heterogeneity, the implication is that they should measure something else than the economic or social phenomenon or phenomena they are supposed to illuminate. There needs to be some distance between the explanans (the thing that explains) and the explanandum (the thing that requires an explanation). This can be problematic as is the case, for instance, with personality indicators. Many trait researchers argue that standard empirical measures of personality are fundamentally unable to explain individual behavior and differences in outcomes (e.g. Bandura 1999: 166; Caprara and Cervone 2000: 16-18, 114-115; Funder 1991: 35-36; Harré 1998: 79-81; Matthews et al. 2009: 3).[20] The reason is that traits are often measured on the basis of the same behavior and/or outcomes they are asked to illuminate. We propose using explanatory capacity as a reference term for dealing with this issue in the context of the study of the economics of preference heterogeneity.

2.4.2. Evaluating Revealed Preferences

Below we evaluate the psychometric quality and usability of RP measures of preference heterogeneity. We thereby focus more on experimental data and less on data from the field. The reason is that field data are mostly available by chance while economic experiments are real measurement devices (cf. Camerer and Fehr 2004; Croson and Gächter 2010; Guala 2008; Plott and Smith 2008: xi). Of course things may change over time, and this also holds for some of our (sub)criteria: data may become increasingly available and new (experimental) techniques may increase the psychometric quality of measures of preference heterogeneity.

Psychometric Quality

Not much is known about the psychometric quality of RP measures, either in general or applied to the measurement of systematic differences in preferences between groups. The absence of such evaluation appears to fit the formal view of preferences (see Section 2.1). If

[20] We can trace discussion of this issue further back to Lewin's (1935 [1931]) discussion of Aristotelian and Galilean modes of causal explanation in psychology. In the distinction he draws between the two, Aristotelian explanation would be circular: "the Aristotelian habit" is to consider "the abstractly defined classes as the essential nature of the particular object and hence as an explanation of its behavior" (Lewin 1935 [1931]: 15) (see Caprara and Cervone 2000: 16).

preferences are both defined and measured as choice (cf. Sen 1977: 325), there is no real need to think about the psychometric quality of preference measures.[21] The evidence on the reliability and validity of RP measures of preferences that is available subsequently is indirect and anecdotal. It follows mostly from studies assessing the sensitivity (or lack thereof) of measured preferences to unobtrusive cues and specifics of the research design. The discussion below is the first of its kind, but unfortunately rather sketchy. We discern three lines of evidence, which all relate to the idea that "there is no such thing as purely internal consistency of choice" (Sen 1993: 499).

The first concerns framing and other contextual effects. Haley and Fessler (2005) provide an illustrative example. They report that in anonymous dictator games played with stylized eyespots present—a cue that does not in anyway affect potentially relevant economic parameters—average offers of a $10 pie increased by 83 cents relative to the situation without such eyespots. Other experiments show the effect of framing, for example, when prospects in a choice experiment receive a positive or a negative frame (Tversky and Kahneman 1981), or when participants in the ultimatum game are labeled buyers and sellers respectively (Hoffman et al. 1994). Results like these tell us something about the reliability (and validity) of RP measures of preferences, specifically that many factors may introduce noise into these measures.

The second line of evidence comes from the large literature on preference reversals (see Tversky and Thaler 1990) and poses a more direct challenge to the reliability and validity of RP preferences. Lichtenstein and Slovic (1971, 1973) pioneered some of the earliest experiments on the preference reversal phenomenon. Using lottery gambles they found that respondents may prefer a prospect (say, Prospect A) over another prospect (say Prospect B) but may actually value the second prospect higher than the first prospect as measured by their willingness to pay (or willingness to sell). Grether and Plott (1979) set out to attribute earlier evidence on reversals in revealed preferences to flawed experimental designs. Improving the design of the experiments did not overturn previous findings, however. Even in experiments completely compatible with economic theory preferences may differ depending on whether they are cast in terms of choices (picking either Prospect A or B) or prices (willingness to pay for either prospect). Theoretically, choices and prices are equivalent measures of preferences,

[21] Several studies in experimental psychology have sought to examine the link between preferences and choice assessing, amongst others, the extent to which choices affect preferences rather than the other way around. See, for example, Savani et al. (2008) and the comment of Sagarin and Skowronski (2009) on Chen (2008) and the subsequent reply by Chen and Risen (2009).

and without saying which one is better, it is clear that RP measures are not free from measurement error.

The most systematic evidence on the reliability and construct validity of RP measures comes from related work on the stability of preferences, specifically those concerning risk. This work finds that it is not uncommon for respondents' preferences to vary with the elicitation procedure used, even to the extent that individuals who are risk-averse under one procedure turn risk-loving in another (e.g. Berg et al. 2005; Isaac and James 2000; Anderson and Mellor 2009). Through Andersen, Harrison et al.'s (2008b) examination of the temporal stability of risk preferences we even get an insight into the test-retest correlation of (risk) preferences. Their analysis shows that the correlation between preferences measured at a 17-month interval varies between 0.34 and 0.58 (Andersen, Harrison et al. 2008b: 1102-1103). This compares with reliability ratios of income and education measures in surveys of about 0.90 (Krueger and Schkade 2008: 1836).

These three lines of evidence show that RP measurement is far from tautological. Many factors appear to affect preferences measured using the RP approach. The result is a non-negligible amount of measurement error. Beyond that we may ask which procedure is best, i.e. which procedure renders the most correct preferences. In the context of measuring preference heterogeneity this question is not very relevant, however. The question is not whether preferences are measured reliably and validly, but whether *differences* in preferences are measured reliably and validly. The effects of a particular elicitation procedure or subtle cues like eyespots need not matter if the purpose is to measure preference heterogeneity, but they may matter a lot if the aim is to value public goods or to predict demand for a new product. If experimental circumstances are the same for all groups, framing or context effects only become problematic when they differentially affect the groups whose preferences are assessed. Measured group variation in preferences may then over- or underrate "true" differences in preferences between the groups. On this specific count, direct evidence is lacking. Andersen, Harrison et al.'s (2008b) findings on the test-retest correlation of (risk) preferences do suggest, however, that also preference heterogeneity is measured with a substantial amount of noise.

Usability

Availability, Scope, and Resolution

Assessing preference heterogeneity using the RP method has a relatively short history. Correspondingly, there are no large-scale datasets available. It may not be prohibitively difficult (expensive) to extend some on-going cross-cultural research projects—the WVS springs to mind—to include RP preference measures (i.e. behavioral experiments), but until this has been done the availability of comparable RP data is limited. Given the inherent difficulties of organizing a comprehensive cross-cultural research project, a meta-analysis of existing studies seems the most practical alternative. Given the sensitivity of RP measures to variations in experimental set-up this is far from ideal, however.

Concerning scope, some of the preferences economists and other social scientists are interested in studying have to do with non-existing or abstract consumption bundles. The RP approach does not seem well-equipped to operationalize and measure this type of preferences. How, for example, can researchers use experiments to give empirical content to preferences towards freedom, democracy, fertility, or entrepreneurship? It does not seem possible to devise experiments that adequately measure these particular preferences. And what about other macro conditions people are likely to face such as law and order, or respect for the elderly? Researchers constantly devise new experiments, extending the range of preferences that can be studied using the RP approach. Nevertheless, there does not seem to be a straightforward way to apply the RP approach to measure and assess preferences concerning such important issues as freedom or democracy, which also likely play a central role in shaping social and economic life.

This point is related to an issue that came up when contrasting the RP and the SP approach (Section 2.3). One factor distinguishing the approaches is that the RP approach makes choices consequential, mostly by allowing experimental subjects to play for money, whereas the SP approach does not. Preferences for many states of affairs cannot be made consequential, however. This holds for many macro-phenomena but, of course, also for individual decisions such as those concerning fertility or how much time to spend working versus the amount of leisure to consume. Hence, measuring preferences of these types in an experimental setting is more likely to be an application of the SP approach than of the RP approach.

Practically, there is of course also the problem that particular preferences may be measured experimentally but that so far this has not yet been done. In this sense, the scope of

EP measures of preference heterogeneity, like the scope of the two other two measures, is also limited by availability.

The resolution of the RP approach on the other hand is high. Data are at the individual level and each observation provides direct information about a specific individual's preferences.

Explanatory Capacity

RP theory is typically applied in studies that aim to predict behavior (e.g. McFadden 1974). Much of the work on the economics of preference heterogeneity on the other hand is aimed at understanding and ultimately explaining economic and social phenomena.[22] Related to this goal difference we find that RP research provides ample evidence of systematic variation in preferences, but that RP measures have limited explanatory capacity.

The problem is that with RP measures of preference heterogeneity, the explanans, which should do the explaining, and the explanandum are close to indistinguishable. On the one hand, observed choice behavior is the essential input for RP analysis. On the other hand, the behavior principally observed to assess the breadth and scope of an individual's preferences is itself an outcome. If the explanans and the explanandum are not separable, the result is, as pointed out by Hanson (1958), Nozick (1981), and others, a tautological argument. Sen (1973: 243) puts it succinctly where notes that "explaining behavior without reference to anything other than behavior is pure rhetoric."[23]

The problem of distinguishing between explanans and explanandum seems most severe for field studies of heterogeneous revealed preferences. It is not clear, for example, what the simple region-of-origin effects in Ichino and Maggi's (2000) study of shirking in Italian banks mean. Similarly, Fernández and Fogli's (2006) finding that fertility of U.S. migrants correlates with home-country fertility show that preference heterogeneity is important but leaves the actual difference in fertility rates unexplained. At the worst, studies like this are at risk of reversing the order of analysis needed for explanation. This is a distinct possibility when preference heterogeneity is measured in terms of a residual, the part of behavior that is left unexplained when all other relevant factors are controlled for.

---

[22] Hempel (1965: 354 and onwards) is the classic treatment of prediction and explanation. As indicated in the previous section, Friedman (1953) argues that economic theory should be judged by its ability to predict.

[23] Sen (1973: 243) makes this point counter to the claim by early RP pioneer Little (1949: 97) that "[i]f an individual's behaviour is consistent, then it must be possible to explain that behaviour without reference to anything other than behaviour."

Importantly, the limited explanatory capacity of RP measures of preference heterogeneity does not preclude their use in other scientific endeavors. Quite the opposite, they can have substantial value in studies that aim to predict economic outcomes. While we cannot explain the color of a table by the color of its atoms, the color of its atoms may very well predict the color of the table, to paraphrase Hanson (1958: 121). There may even be a trade-off between prediction and explanation. Funder (1991), in particular, finds that the predictive power of measures of preferences or some related mental entities may be improved by narrowing the construct, as when the construct "social skill at parties" is measured by observing social skill at parties, and subsequently used to predict social skill at a party (Funder 1991: 35). Global constructs on the other hand are able to relate specific observed behaviors or socioeconomic phenomena to a more complex and general pattern of behavior, which, according to Funder (1991: 36), is what explanation is all about.[24] Simply put, a move from the specific to the general enhances explanatory capacity whilst diminishing predictive power (see also Hechter et al. 2005).[25]

Importantly, explanatory capacity here is not the same as explanatory power. Explanatory capacity is a fundamentally theoretical property of psychometric measures and derives from the nature of scientific explanation. Whether measures with high explanatory capacity, i.e. measures clearly distinguishable from the explanandum, actually have explanatory power needs to be proven in practice. Findings like those of Hechter et al. (2005) suggest that an increase in explanatory capacity lowers the correlation between measures of preference heterogeneity and socioeconomic outcomes, regardless of the actual relation between the two. The reason is simply that distancing the explanans from the explanandum introduces noise in the statistical relation.

The issue of prediction versus explanation resurfaces briefly in the conclusion to this chapter (Section 2.6) where we discuss other possible uses of RP measures of preference heterogeneity.

---

[24] This idea matches very well Friedman's (1953: 4) notion of what positive economics should do: "Its task is to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances."

[25] Related is the precision-prediction paradox, which Smits et al. (2003: 3) trace back to classic handbooks on measurement in psychology by Cronbach and Gleser (1965: 136-137), Gulliksen (1950: 380-381), and Lord and Novick (1968: 332) (references as they appear in Smits et al. 2003). High reliability requires high correlations between the different items making up an indicator but this lowers the contribution each item makes to the variance in indicator scores and thereby the indicator's power to correlate with outside phenomena.

2.4.3. Evaluating Stated Preferences

Psychometric Quality

Skepticism towards the reliability and validity of SP measures is widespread in economics (although the terms reliability and (construct or measurement) validity are not widely used). A priori, many economists find that talk is cheap and that behavior is all that matters. This skepticism has historical roots dating back to at least 60 years ago. McCloskey (1983: 485-486, 514) traces it back to the effort in the late 1930s to ask businesspersons whether they equalized marginal cost to marginal revenue and, particularly, Friedman's (1953: 31) dismissal of this kind of evidence. Manski (2004) sketches a similar picture of widespread skepticism towards SP data and finds it is rooted in Machlup's (1946) dismissal of a similar survey on businesspersons' cost and revenue expectations (see, also, Dominitz and van Soest 2008). Bertrand and Mullainathan (2001) discuss some potential weaknesses of subjective survey data that economists tend to emphasize. They distinguish different sources of biases and general errors (see also Fischhoff 1991). Some of these may apply equally well to the other two approaches, revealed and experienced preferences.

For starters, cognitive biases may occur, meaning that framing and specific wording affect people's answers (cf. Tversky and Kahneman 1981). In addition, respondents may exercise little mental effort so that they do not take in all possible answers offered for a particular questionnaire item. A second issue is that of social desirability bias. Respondents may profess to entertain certain value preferences only to please the interviewer.[26] Thirdly, attitudinal surveys face the problem that they aim to measure a construct that does not actually "exist" (cf. Fischhoff 1991). One possible effect is that for their responses people, unable to answer, simply base their stated values and attitudes on past behavior. Finally, there may be biases due to strategic responses. These are not included in Bertrand and Mullainathan's (2001) overview, but particularly relevant for non-market valuation such as contingent valuation and measures of stated willingness to pay and willingness to accept more generally (cf. Cameron 2008).

Taken together, biases and errors like these imply that survey-based measures of preferences and attitudinal traits have limited precision and accuracy. This could potentially invalidate SP measures of preference heterogeneity. As with RP measures, this occurs when

[26] This type of bias of course is closely related to the observation that in behavioral experiments, next to explicit (monetary) incentives, implicit incentives are plentiful (cf. Loewenstein 1999: F31; Levitt and List 2007).

the biases and errors are not random measurement noise but for some reason vary across the classes for which preferences are measured. A bias means that measurement noise is not random, but only if a certain bias differentially affects the groups whose preferences are compared does it imply that measured group variation in preferences over- or underrates "true" preference heterogeneity.

Outside of economics, broad-ranged evaluations of SP data primarily assess the work of Hofstede (1980, 2001), Inglehart (e.g. 1990, 1997), and other comprehensive culture datasets. Critical appraisals question, inter alia, the overall validity of the frameworks that are applied (e.g. Ailon 2008; Baskerville 2003, 2005; Fiske 2002; Haller 2002; MacIntosh 1998; McSweeney 2002a, 2002b; Taras et al. 2009; Taras et al. 2010; Venaik and Brewer 2010). Hofstede (2001: 73; 2002: 1356; 2010) provides overviews of such critiques (and rebuttals).

A specific, and more detrimental, challenge to the psychometric quality of SP measures of preference heterogeneity is also available in this literature. Several studies examine the extent to which stated values, preferences, attitudes, and the like are sensitive to changing circumstances, generally with damning results. Clarke et al. (1999) and Duch and Taylor (1993), for example, assess the validity of Inglehart's measurement of values in the WVS. Their analyses show that the responses to the questionnaire items involved are very sensitive to short-term economic and political conditions. Based on these findings, they conclude that this survey does not measure deeply ingrained and durable value dispositions but rather salience (see, also, Clarke and Dutt 1991; Clarke et al. 1993; Clarke et al. 1997; Davis 1996; Davis and Davenport 1999; Davis et al. 1999).

Here again the question is whether groups are differentially affected. Evidence on this point comes from Maseland and van Hoorn (2009). They look at the relation between the values and practices measures of national culture as measured by the GLOBE project (House et al. 2004). The negative correlations between values and practices in seven out of nine dimensions (House et al. 2004: 736) suggest to Maseland and van Hoorn (2009) that circumstances have a systematic, biasing effect on stated preferences. They subsequently invoke the principle of diminishing marginal utility to explain these seemingly counter-intuitive negative correlations between values and practices (Brewer and Venaik 2010; Hofstede 2006; Javidan et al. 2006; Taras et al. 2010). The idea is simply that people state that they attach high value to law and order when they live in a state of anarchy and low value when society tightly regulates people's moves. If this explanation is correct, SP measures of preference heterogeneity such as those available from GLOBE, Hofstede, and the WVS actually capture group differences in *marginal* preferences—the preference for increasing

satiation of an objective given current levels of satiation, rather than values or attitudes—the preference for satiating the objective in general (Maseland and van Hoorn 2009). On this interpretation, SP measures would be invalid, failing completely in capturing the construct they purport to capture.

Chapter 3 is devoted to the study of this possible "marginal preferences problem," using an (alleged) democracy paradox of Islam (Bratton 2003; Hoffman 2004; Inglehart and Norris 2003; Jamal and Tessler 2008; Mogahed 2006; Rose 2002; Rowley and Smith 2009) as a case study. Let us already note, however, that the problem, if borne out, implies an inability to explain: variation in stated preferences rather than accounting for differences in (socioeconomic) circumstances between groups would be the result of such differences.

Usability

Availability, Scope, and Resolution

SP measures of preferences score high on all three counts, availability, scope, and resolution. The measures are widely available. What is more, they tend to come in a format that allows for easy assessment of preference heterogeneity, meaning that responses from many thousands of individuals concerning a broad range of issues can be analyzed. Detailed personal information is also available, including respondents' nationality, religious denomination, gender, marital status, and other sociodemographic characteristics. In the appendix to Chapter 4 we review several large-scale data-gathering efforts containing the kind of data just mentioned, most of which are publicly available.

The SP approach offers similarly broad opportunities with regards to the scope of its measures. Preferences can literally concern anything, concrete consumption bundles or abstract, non-existing ones. People may even be asked about past events or historical figures (see Section 2.1). As indicated above, the only limitation on SP measures' scope actually concerns availability. Of course, (cross-cultural) questionnaires like the WVS include only a limited number of items and if a specific stated preference is missing, a comparable survey is not easily organized.

Finally, the resolution of the SP approach is at the same level as that of the RP approach. Each SP measure provides direct information about the preferences of an individual.

Explanatory Capacity

The explanatory capacity of SP measures of preference heterogeneity faces the same threat as RP measures do. As discussed in Section 2.3, the SP approach and the RP approach have a certain resemblance towards each other (e.g. hypothetical vs. non-hypothetical choice experiments; see Figure 2.1). To the extent that the difference between the two is indeed a matter of degree, which could hold for some but definitely not all SP items, SP measures also make it problematic to distinguish between the explanans (the thing that explains) and the explanandum (the thing that requires an explanation). The fact that an individual stated a preference for Good A over Good B may predict the actual buying of Good A and not Good B when given the choice, but does little to explain this single behavioral event.

A possible solution may again be found in the broadening of constructs (Funder 1991: 35-36). A broader pattern of stated preferences may properly explain narrow phenomena and outcomes as elements of set pattern (ibidem). Importantly, frameworks like those of Hofstede or GLOBE offer such broad constructs, reducing cultural differences between countries to four or maximum nine dimensions. On the basis of frameworks like these, researchers can generalize specific socioeconomic differences between countries to broad-based country variation in preferences.

2.4.4. Evaluating Experienced Preferences

Psychometric Quality

The EP approach to measuring preferences has only recently been developed and not much is known about its psychometric quality. The discussion has mostly focused on the measures of SWB that provide the necessary input for the approach. These have met with the same skepticism as SP data (see above). Much doubt exists whether SWB indicators are meaningful in the sense that they convey important and useful information on individuals' well-being. Chapter 5 deals extensively with the question whether SWB ratings are reliable and valid. In general, they appear to have a reasonable level of psychometric quality. The measures can be noisy, however, and have a demonstrated sensitivity to whimsical circumstances like the outcome of a soccer match or the weather (Schwarz and Strack 1999: 62). Practically this means that the accuracy of measuring SWB is improved by increasing the number of observations, which helps eliminate such random influences. The EP approach typically compares large groups so that neither whimsical circumstances nor any other source of measurement error is likely to differentially affect levels of SWB of the groups involved

(cf. Di Tella and MacCulloch 2006: 29; Frey et al. 2009: 320; Krueger and Schkade 2008: 1843). And even when they do, it does not follow that EP measures of preference heterogeneity are invalid.

The EP measures do have a validity problem, however, when the effect of a situational factor like leisure or income on experienced utility does not vary with the preferences associated with group membership but with some other collectively shared factor (cf. Finkelstein et al. 2008: 22). If this identifying assumption were found incorrect, this would pose a most fundamental challenge to the validity of any measure of preference heterogeneity. For none of the approaches is it possible to test the validity of the identifying assumption directly. That is, even when we observe systematic differences in experimental choices, stated preferences, or the shape and content of the experienced utility function across groups, we cannot be sure that these differences are due to actual differences in preferences rather than to some other factor that varies systematically across the groups involved. Fundamentally, none of the approaches establishes a direct causal link as to why we observe differences in experimental choices, variation in stated preferences, or differentially shaped experienced utility functions. Preference heterogeneity is the most logical explanation but due to the quasi-experimental nature of the approaches, this is not causally demonstrated (cf. Note 11; see, also, van de Vijver and Leung 1997: 2-3).

It may subsequently be argued that the EP approach requires an overly strong identifying assumption (more so than the other two approaches to measuring preference heterogeneity). Finkelstein et al. (2008: 22-28) bring some evidence to bear on this issue. They identify a priori possible sources of systematic variation in the empirical structure of happiness other than group regularities in the content and shape of the experienced utility function. Specifically they examine whether their finding that the marginal utility of consumption declines with worsening health is, somehow, driven by noise in their SWB indicator that varies with income or health (however unlikely). Using, amongst others, different (experienced) utility proxies, they are able to show that it does not.

Finkelstein et al.'s (2008) assessment of the robustness of measured heterogeneity in experienced preferences does not explicitly address whether the measures are reliable and valid. It is the only evidence available, however. To fill this gap in the EP literature, in Chapter 5 we conduct an empirical assessment of the psychometric quality of the EP approach, for which we expand the basic method applied by Finkelstein et al. (2008).

Usability

Availability, Scope, and Resolution

EP measures of preference heterogeneity are not readily available. The data required to measure experienced preferences and heterogeneity therein, on the other hand, are. Many of the surveys containing comprehensive data on values also contain data on SWB (see Chapter 4 for a further discussion). This makes EP measures relatively easy to obtain. Next to SWB data, the EP approach also requires data on potential determinants of SWB, however. These can come from sociodemographic details also included in the survey, such as marital status, individual income, or employment status. Data on aggregate-level determinants—inflation, inequality, et cetera—can come from many sources. For the EP approach to work, these latter data need to be matched to the available survey data on individual SWB. Nearly always, the basis for this matching will be the individual's country of residence. The reason is that most aggregate variables concern differences between countries.

The need for matching notwithstanding, the EP approach offers great flexibility (cf. Di Tella and MacCulloch 2006: 40). In principle, anything that can be measured can, after matching, be related to SWB to see how it is valued by different groups. The SWB effect can be positive or negative, reflecting how much people like or dislike a certain state of affairs. This flexibility means the EP approach has a broad scope and is able to measure many different types of preferences. The EP approach could run into problems, however, when it is asked to measure risk preferences and time preferences. Risk may be valued indirectly by looking at the happiness effect of volatility in outcomes but this is far from ideal. We see no straightforward application of the EP approach to measure time preferences. The scope of the approach is further limited by its need for variation. As mentioned in Section 2.2, estimating the shape of the average experienced utility function for the members of a group requires that there is individual variation in the determinants of SWB. If variables vary only little, either across individuals (e.g. marital status, income) or within a group over time (e.g. inflation, inequality) measurement of preference heterogeneity will be inaccurate or even become impossible.

The variation requirement also has implications for the resolution of EP measures of preference heterogeneity. Estimation across individuals that differ in the consumption of the objective that is being valued implies that EP measures can only be obtained at the group level. Within an individual there is unlikely to be enough variation in the relevant variables to get a reasonable estimate for a person-specific experienced utility function.

Explanatory Capacity

How distinct are EP measures of preference heterogeneity from the economic or social phenomenon or phenomena they may be asked to illuminate? Clearly, the happiness valuation of a certain state of affairs is something very different from this state of affairs. This detachment of explanans and explanandum implies that EP measures of preference heterogeneity have considerable capacity to account for differences in socioeconomic outcomes. There is a caveat, however. Next to data on experienced utility, the EP approach also requires data on outcomes to measure experienced preferences and group variation therein. Implication is that whilst explanans and explanandum are separated, they are not completely detached in the EP approach to the study of the economics of preference heterogeneity either.

2.5. SCORING APPROACHES AND MEASURES OF PREFERENCE HETEROGENEITY

The previous section has evaluated the three approaches to assessing preference heterogeneity using two core criteria and several sub-criteria. It took a psychometric perspective but also considered the application of measures of heterogeneous preferences in the study of the economics of preference heterogeneity. This short section adds a comparative perspective, evaluating the approaches and their measures of preference heterogeneity relative to each other.

Our approach is simple, scoring the approaches from best (●●●●●) to worst (●). A highest score does not indicate that a particular approach or measure is perfect, without limitations, but that it is the best available. Similarly, we look only at current performance and do not consider potential improvements, for example on the availability sub-criterion (see, however, the concluding section). Where detailed information is lacking, we score the approach in a range of most likely actual scores, again relative to the other approaches (e.g. ● - ●●●). Of course this will always involve some judgment, not least when information is limited. We include all criteria, thus psychometric quality, and usability as divided in the sub-criteria availability, scope, resolution, and explanatory capacity. Table 2.1 summarizes our comparative evaluation of the approaches to assessing preference heterogeneity. A short but more detailed discussion follows. The final section concludes with a discussion of the usefulness of the RP, SP and EP approaches in different types of economic analyses.

To start with the first core criterion, the psychometric quality of measured preference heterogeneity, so far there are no clear-cut cases. All approaches to measuring preferences are subject to measurement error. To the extent that this error is not random noise, it can introduce structural biases in numerical measures of preferences and, more relevant, in measures of preference heterogeneity. As discussed, the latter happens when the error, for some reason, differs between the groups for which preferences are assessed. Overall, SP measures very likely have the poorest psychometric quality. We withhold final judgment until after the detailed case study of the marginal preferences problem in the next chapter, however. Hence, for now, we score the psychometric quality of SP measures of preference heterogeneity on a range. Due to the lack of information, we do the same for the RP and EP approach. Siding with the idea that behavior tends to be more meaningful than a self-report, we deem it likely that the RP approach performs better than the EP approach, which in turn is likely to have better psychometric quality than the SP approach.

Table 2.1: A Comparative Evaluation of the RP, SP and EP Approaches and Measures.

Approach	Revealed Preferences	Stated Preferences	Experienced Preferences
Psychometric quality	● - ●●●●●	● - ●●●	● - ●●●●
Usability			
Availability	●	●●●●●	●●●●
Scope	●	●●●●●	●●●●
Resolution	●●●●●	●●●●●	●
Explanatory capacity	●	● - ●●●●	●●●●●

Turning to availability, we find the SP approach slightly outperforms the EP approach, whilst both fare better than the RP approach. This is all due to the availability of large-scale datasets like the WVS, which form valuable input for these two approaches but not for the RP approach (at least not yet).

Concerning scope, the evaluation of the RP approach finds that its underlying measurement methodology cannot be used to give empirical content to certain economically relevant types of (heterogeneous) preferences, the preference for democracy is an important case in point. In contrast, researchers can use the SP and the EP approaches to measure virtually any preference imaginable. This is most obvious in the case of stated preferences but because the EP approach requires minimum structure it similarly has much more potential for operationalizing and measuring economically relevant preferences than does the RP approach.

We can be brief about resolution. The RP and SP approach can be applied at any level the researcher is interested in, whereas the EP approach can only be applied at the group level. This is a disadvantage, although economic analyses nearly always involve some form of aggregation (see Chapter 1). The resolution sub-criterion therefore does not have much practical relevance for the study of preference heterogeneity.

Finally, the performance of the RP approach stands out negatively in its explanatory capacity. In short, the RP approach runs into fundamental difficulties when used to empirically explain economic disparities on the basis of differences in preferences. As discussed by Hanson (1958) and Nozick (1981), amongst others, explanation requires that the explanans (the thing that explains) can be distinguished from the explanandum (the thing that requires an explanation); and if choice-based measures of preference heterogeneity are used in empirical analyses aiming to illuminate socioeconomic differences it does not seem possible to separate the two. The same problem of distinguishing between the explanans and the explanandum may plague SP analysis. To the extent that the SP approach shares the RP approach's focus on choice, it may similarly suffer a circularity problem, thwarting the explanatory capacity of SP measures of preference heterogeneity. The capacity to explain is highest for the EP approach, though we should emphasize that methodologically it is not perfect either.

2.6. CONCLUSION: MEASURING PREFERENCES FOR THE STUDY OF THE ECONOMICS OF PREFERENCE HETEROGENEITY

Our general conclusion on the three approaches is that there is not one best approach to measuring preferences and assessing preference heterogeneity, either theoretically or practically. The reasons for this are manifold and interrelated. Bottom line is that the RP, SP and EP approaches all have strengths and weaknesses, most of which tend to be specific to the research question the measures are called upon to address. Furthermore, they all face trade-offs, where a lower score on one (sub-)criterion may be exchanged for a higher score on some other criterion, as when reliability or validity are traded off for scope (a comprehensive reflection of group level differences in preferences for many possible objectives). The research question ultimately provides the most relevant guidance for choosing between such features of the empirical measures involved.

The difficulty of passing a final judgment is further aggravated by the lack of information on the psychometric quality of the three types of measures of preference heterogeneity, and by the fact that some of the features of the measures and approaches are not static. We find, for example, that RP measures of preference heterogeneity are not widely available, whereas the SP and EP measures are, or at least can be easily obtained. Availability is hardly static, however, and over time, large-scale research projects may make different types of measures of preference heterogeneity increasingly available. Note, however, that such undertakings involve important pragmatic considerations, not least of which is the cost of collecting the data.

As is, we are only able to offer a preliminary conclusion concerning the usefulness of the three approaches, RP, SP and EP, for the study of the economics of preference heterogeneity. Partly by elimination, we find the EP approach has the most potential. Based on the information gathered in this chapter, we dismiss the RP approach altogether. The reason is that RP measures of preference heterogeneity, though possibly doing well in the area of reliability and validity, lack the (theoretical) capacity to explain, which is not compensated for by outstanding performance on the other usability criteria (particularly scope and availability). SP measures, on the other hand, may be able to compensate their possible lack of psychometric quality, but a more precise ruling requires further insight in the marginal preferences problem and its severity. We turn to this in the next chapter. Based on the empirical analysis in that chapter we can update the summary evaluation of the SP approach presented in Table 2.1 and likely draw stronger conclusions concerning the overall usefulness of the SP approach (relative to that of the EP approach). Similarly, very little is known about the psychometric quality of EP measures of preference heterogeneity but the approach does quite well on availability, scope, and explanatory capacity. The second part of this thesis (Chapters 4 and 5) is entirely devoted to assessing the psychometrics of indicators of SWB and their use in the measurement of preferences and heterogeneity therein. If the promise of usability is backed up by a reasonable degree of reliability and validity, the relatively novel EP approach may be developed into a valuable tool for studying the economics of preference heterogeneity.

To be sure, this latter conclusion does not imply that the other two approaches have no use whatsoever in the study of preferences. RP measures of preference heterogeneity in particular have valuable scientific uses. Firstly, they may be useful for giving a systematic description of differences in preferences between selected groups such as men and women or

people from different societies. Secondly, they have value if the aim is to predict—as opposed to explain—variation in specific socioeconomic outcomes between groups.

Notwithstanding, in light of the study of the economics preference heterogeneity the EP approach's potential is very appealing. The remainder of this thesis is concerned with evaluating the approach further (Part II, Chapters 4 and 5) and with applying it to some long-standing traditions of researching preference heterogeneity in social science (Part III, Chapters 6 and 7). First, however, the next chapter gives a detailed assessment of the marginal preferences problem in SP data.

Chapter 3

An Analysis of the Marginality of Stated Preference Measures with an Application to the Democracy Paradox (of Islam)•

• The parts in this chapter containing the background to the democracy paradox of Islam and the empirical analysis are taken from joint work with Robbert Maseland, published as Maseland, R., and A. van Hoorn. 2010. Why Muslims Like Democracy Yet Have So Little of It. *Public Choice*, doi:10.1007/s11127-010-9642-5 (Maseland and van Hoorn 2010b).

3.1. INTRODUCTION

This chapter expands on the idea, mentioned in the previous chapter, that SP data are so sensitive to respondent circumstances that they mistake marginal preferences for deep-rooted attitudinal traits. We present theoretical background to the marginal preferences argument and empirically investigate it by applying its logic to the Muslim democracy paradox that has puzzled many researchers.

Over the last decade numerous scholars have been drawn to the so-called democracy paradox of Islam: the fact that Muslim-majority countries tend to be less democratic, while both individual Muslims and individuals in Muslim-majority countries tend to have a much more positive attitude towards democracy than other people do (Bratton 2003; Hoffman 2004; Inglehart and Norris 2003; Jamal and Tessler 2008; Mogahed 2006; Rose 2002; Rowley and Smith 2009). To these authors, this gap between democratic attitudes and institutions in the Islamic world is paradoxical. It has to be explained by other factors; gender equality (Inglehart and Norris 2003) and religious freedom (Rowley and Smith 2009) are suggested candidates.

Taking the measurement of heterogeneous preferences as our starting point, we investigate the possibility that the paradox is not a paradox after all, but rather reflects the limited psychometric quality, specifically validity, of stated preference (SP) measures of preference heterogeneity. First, as we discussed in the previous chapter, the sensitivity of stated preferences to circumstances is well-established (Clarke et al. 1999; Duch and Taylor 1993). Second, and more importantly, the marginal preferences argument (Maseland and van Hoorn 2009) actually suggests that positive stated attitudes tend to go together with low levels of satiation—not with high levels. On this count, looking for omitted variables to explain the democracy puzzle of Islam is misguided as there is no puzzle. Rather, the alleged Muslim democracy paradox can be understood in terms of the basic economic principle of diminishing marginal utility: if something is scarce, we value it more. Muslim-majority countries tend to be less democratic, so its inhabitants have a craving for more democracy. There is nothing paradoxical about this; it is basic microeconomics.

For our empirical analysis we use data from the World Values Survey (European Values Study Group and World Values Survey Association 2006; World Values Association 2009), which we supplement with democracy data from the Polity IV project (Marshall 2008). These are the same data as used in most previous research on the Islam-democracy nexus. Our main analyses are based on a sample of 68 countries and more than 100,000 individuals.

Results show that the negative relation between democratic attitudes and actual levels of democracy is not unique to Muslim countries. Using five different measures of attitude towards democracy, we find that the negative relation can be generalized from Muslim countries to all countries in our sample. This evidence is consistent with SP data suffering a marginal preferences problem. It appears that SP measures of preference heterogeneity do not capture attitudes, the preference for satiating an objective in general but marginal preferences, the preference for increasing satiation of an objective given current levels of satiation. Whereas differences in attitudes may explain differences in outcomes such as levels of democracy, marginal preferences are *driven by* rather than *drivers of* differences in environments.

Our findings have fundamental implications for the validity of SP measures and their use in the study of the economics of preference heterogeneity, as we elaborate at the end of this chapter. Bottom line is that our empirical evidence suggests that SP measures of preference heterogeneity are invalid, systematically misrepresenting group differences in preferences, and that they therefore cannot be used to study the economics of preference heterogeneity. We incorporate this conclusion in an update to the results of the evaluation of the usefulness of the SP approach in the study of the economics of preference heterogeneity in Chapter 2.

The structure of this chapter is as follows. We start by elaborating the marginal preferences problem in Section 3.2. Section 3.3 briefly considers the economics literature on religion and discusses the determinants of democracy and democratic attitudes, specifically the regularly observed Muslim democracy paradox. We present our (empirical) explanation for the democracy paradox of Islam in Section 3.4. We derive two competing hypotheses by which to evaluate our case and test these using multilevel modeling. The results strongly support the hypothesis that SP data suffer a marginal preferences problem that accounts for what superficially looks like a democracy paradox. Section 3.5 presents concluding remarks including a brief discussion of the far-reaching implications of the marginal preferences issue for the validity of SP measures of heterogeneous preferences and the usefulness of such measures in the study of the economics of preference heterogeneity.

Chapter 3 concludes Part I of this thesis. Directly following this chapter, we present a brief summary of our findings thus far. The focus thereby is on the comparative usefulness of the three approaches to measuring preferences (RP, SP, and EP) for the study of the economics of preference heterogeneity.

3.2 VALUES VERSUS MARGINAL PREFERENCES

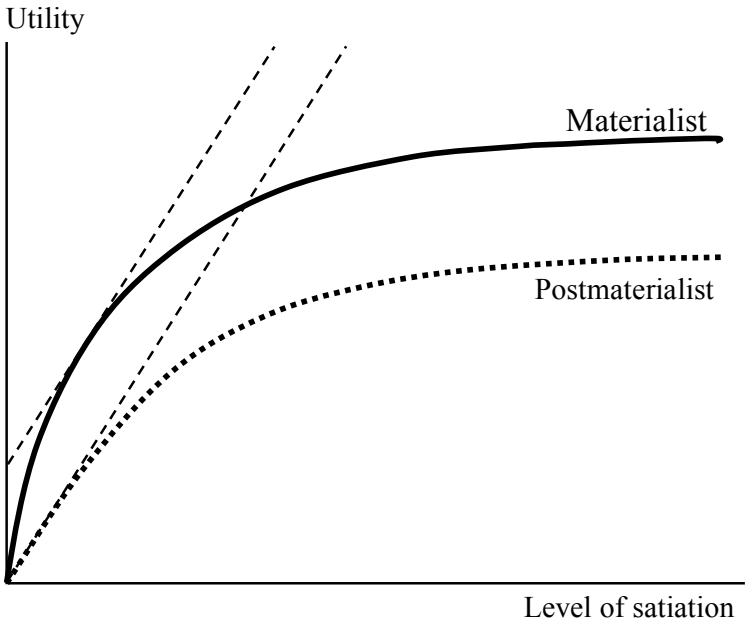
The idea behind the marginal preferences problem is surprisingly simple. Take a questionnaire item asking respondents whether they agree or disagree that work not leisure is what makes life worth living. Judging this item by its content, we may well be led to find that it elicits information about respondents' attitude towards work as compared with leisure. The corresponding interpretation of answers given would be that respondents expressing agreement with the statement value work more highly than leisure and have stronger work ethic than those who disagree. The marginal preference argument, by contrast, finds that things are not always what they seem. The fundamental proposition is that expecting respondents to indicate the importance attached to certain objectives completely independent of the extent to which such objectives are currently satiated is placing too strong cognitive demands on people (Maseland and van Hoorn 2010a).

This proposition incorporates two related literatures. The first of these, discussed in Chapter 2, systematically demonstrates that measures of attitudes and values produced by survey instruments may fluctuate heavily with changing circumstances (e.g. Clarke and Dutt 1991; Clarke et al. 1999; Davis and Davenport 1999; Davis et al. 1999; Duch and Taylor 1993). The second body of research may hardly be considered systematic, consisting of a hodge-podge of questionnaire-based studies of the economics of preference heterogeneity reporting "intriguing" and "counterintuitive" findings. A first, prominent example of such work is the GLOBE project (House et al. 2004). This project set out to measure cultural differences between countries along nine dimensions (incorporating and the pioneering work extending of Hofstede 1980, 2001). Covering 62 societies it is the most comprehensive cross-cultural study thus far. Much to their own surprise, the people involved with this project found strong evidence that the value attached to objectives and the degree to which these objectives are satiated in a society are negatively correlated (House et al. 2004: 729-730, 736; Javidan et al. 2006; see also Brewer and Venaik 2010; Maseland and van Hoorn 2009; Taras et al. 2010). In similar fashion, throughout his work, Ronald Inglehart (e.g. 1997) has consistently shown that those societies that attach the most value to materialist objectives, which concern material needs such as security, shelter and sustenance, are poorer than societies that care relatively less about material goals and favor postmaterialist objectives such as autonomy and self-expression instead.

The latter is not usually regarded a paradox, however. On the contrary, Inglehart (1997) has a perfectly plausible explanation for the value change he observes when societies

grow richer. The richer a society becomes, the less importance people attach to growing richer still. Wealth, like most other objectives, faces diminishing marginal utility (see also Inglehart 1987, 1990). And it is the same principle of diminishing marginal utility that lies at the heart of the marginal preferences argument.[27]

Figure 3.1, Choice between Materialist and Postmaterialist Objectives.



To illustrate, Figure 3.1, which is similar to Figure 2.2, depicts a choice between materialist and postmaterialist objectives. We use this figure to elaborate how the principle of marginal utility applies to SP data, specifically their sensitivity to respondents' circumstances, and undermines their validity. The chief thing to note is that the eventual preference for postmaterialism in general is independent of the fact that, overall, society values materialism more. Society would still choose materialism over postmaterialism any time when satiation levels are equal. In other words, the attitude towards materialism does not change; it is only the preference for more materialism on top of the current endowment that may change. This distinction is critically important because only the first corresponds to what the literature identifies as cultural values or attitudes (e.g. Hofstede 1980; House et al. 2004; Inglehart

[27] Inglehart's postmaterialist value shift thesis is actually more sophisticated than this, arguing that shifts in culture come about through generation cohorts, whose values are shaped by material circumstances during childhood, so-called "socialization." The generation argument, however, is about the way value change takes place. The basic mechanism behind it is that people value less what they have more off.

1990, 1997),[28] and only the first can function as an autonomous factor in determining behavior and outcomes. The latter, which we may dub the marginal preference for materialism, is determined by outcomes rather than the other way around (i.e. determining them).

Turning back to the marginal preferences problem we can state it succinctly as people's professed values or attitudes systematically responding to levels of satiation so as to signify marginal preferences—the preference for increasing satiation of an objective given current levels of satiation, instead of deep-rooted cultural traits associated with the preference for satiating an objective in general (or weights in the utility function). From a comparative point of view, this interpretation of values survey scores implies that societies that emphasize postmaterialist objectives the most are precisely those that have the highest satiation of materialist relative to postmaterialist objectives. As such, this negative relation may seem paradoxical but only if one expects stated preferences to be explanatory variables determining outcomes. If, in contrast, one recognizes how people's stated preferences respond to differences in circumstances according to the principle of diminishing marginal utility, a negative relation is exactly what one would expect.

We apply this logic to account for Islam's democracy paradox in Section 3.4 below. First, however, we give further background to the study of religion in economics and statistical data on this alleged paradox.

3.3. BACKGROUND ON RELIGION IN ECONOMICS AND DEMOCRACY IN MUSLIM COUNTRIES

3.3.1. Religion in Economics

Interest among economists in religion and religious behavior has increased in recent years. Starting with Laurence Iannaccone's seminal contributions (Iannaccone 1988, 1992a, 1992b), an expanding literature has developed that applies economic insights to religious behavior. The objective of this body of research is to explain religious behavior with the common tools of rational choice analysis. Some examples of this approach are the studies by Epstein and Gang (2007), Iannaccone and Berman (2006) and Kurrild-Klitgaard et al. (2006), all of which

[28] See the definitions given in the second section of Chapter 2.

explain the occurrence of religious fundamentalism and terrorism in terms of a rational choice framework. This work shows that lack of freedom (Kurrild-Klitgaard et al. 2006), lack of effectively functioning states and markets (Iannaccone and Berman 2006), and the presence of competition between religious groups (Epstein and Gang 2007) are factors conducive to religious extremism and militancy.

A second class of literature focuses on religious variation in (economic) beliefs and attitudes. This work follows the path laid out by Weber's (1904/1905 [1930]) examination of a unique Protestant ethic as underlying the rise of the modern capitalist system. Research of this type fits economics' increasing interest in preference heterogeneity and has gained popularity over the past few years. Prominent examples are the studies by Guiso et al. (2003, 2006) and McCleary and Barro (2006), which we briefly discussed in the previous chapter. This work is rather general, covering many types of beliefs and attitudes and not typically linking them to one or more specific denominations. Other research is less broad, focusing more on attitudes associated with certain denominations, specifically Protestantism. Torgler and Schaltegger (2009), for example, report that the general link between the religious belief in hell and work ethic found by McCleary and Barro (2006) is stronger in societies with a high share of Protestants whereas Catholicism has no statistically significant effect. In a follow-up study, the same authors examine religious variation in value attached to work for a sample of 17,000 individuals from 16 European countries (Schaltegger and Torgler 2010). Results point to education as a channel linking Protestantism to work ethic. For Protestants, a higher education level is associated with stronger agreement with the statement that work should come first even if it means less spare time. Finally, Arruñada (2010) finds that Catholics and Protestants do not differ in work ethic, but that Protestants have a stronger "social ethic" as reflected in a greater trust in institutions, less tolerance for tax fraud, and lower willingness to cover up for friends. Overviews of empirical work on the Weber thesis can be found in Iannaccone (1998: 1474-1478), Furnham (1984, 1990), and Jones (1997). We discuss Weber's work and some of the empirical research it has spawned in more detail in Chapter 6.

A final class of literature examines the consequences of religious beliefs and ideology for political and also economic outcomes. In contrast to the rational choice approach to religion, this literature takes religious beliefs and ideology as exogenous inputs to political and economic behavior rather than seeing them as products of rational decision making. This approach also fits economics' increasing interest in preference heterogeneity and has been given a lift by Huntington's famous clash of civilizations thesis (Huntington 1993, 1996),

which boosted interest in the political and economic effects of religious values. Some research in this area is based on historical case studies (Kuran 2003, 2009), though most makes use of quantitative analyses relating political and economic outcomes to religious and cultural beliefs and values (e.g. Barro and McCleary 2003; McCleary and Barro 2006; Guiso et al. 2006). As an early example of this trend, Glahe and Vorhies (1989) show that Judeo-Christian liberal values are conducive to economic development. As part of this literature, several studies have focused on the impact of the Islamic faith on democracy (Karatnycky 2002; Midlarsky 1998; Rowley and Smith 2009; see also Inglehart et al. 2002). This research has produced some supposedly paradoxical results (Bratton 2003; Hoffman 2004; Inglehart and Norris 2003; Jamal and Tessler 2008; Mogahed 2006; Rose 2002; Rowley and Smith 2009).

3.3.2. The Democracy Paradox of Islam

The democracy paradox of Islam, according to Rowley and Smith (2009), is the fact that Muslim-majority countries tend to be significantly less democratic than other societies, although their citizens and, in fact, individual Muslims everywhere appear to entertain a more positive attitude towards democracy than others do.

The first part of this paradox is readily accepted. The determinants of democracy have been well researched, a primary driver being economic prosperity (Lipset 1959). Part of this relation between economic prosperity and the level of democracy runs through its effect on education and the size of the middle class (Evans and Rose 2007; Glaeser et al. 2007; Hadenius 1992; Kamens 1988; Lipset 1959; Shafiq 2010). However, with measures of education and middle class controlled for, income continues to have an independent effect as well (Barro 1999). Further factors found to affect the prevalence of democracy are natural circumstances such as country size and insularity (Anckar 2002; Barro 1999; Clague et al. 2001; Hadenius 1992), although the theoretical mechanism behind these relationships is rather obscure. In addition, income inequality, the urbanization rate, and ethnic heterogeneity have been argued to influence democracy. Statistically observable effects of these factors are relatively weak, however (Barro 1999). On top of such general determinants of democracy, researchers have called attention to historical particularities such as colonial legacy (Anckar 2002; Clague et al. 2001; Huntington 1984; Weiner 1987), religious background, and cultural traditions (Clague et al. 2001; Huntington 1993; Inglehart 1988). In particular, Islam consistently shows up as a significant, negative factor in many empirical studies (Barro 1999; Fish 2002; Karatnycky 2002; Midlarsky 1998; Ross 2001; Rowley and Smith 2009).

Several explanations have been proposed to explain the latter empirical regularity. These range from the absence of separation of church and state in Islam (Huntington 1993, 1996; see also Minkenberg 2007), through gender inequality (Fish 2002; Inglehart and Norris 2003), to endowments of oil (Barro 1999; Ross 2001). As it turns out, material conditions such as the availability of oil and the associated inequality explain some of the observed variation in democracy levels, but do not take away the significantly negative relation between a Muslim background and democratic institutions (Barro 1999; Fish 2002; Rowley and Smith 2009). The answer thus seems to lie in beliefs and attitudes associated with Islam itself.

Yet, by now, a vast literature has emerged showing that Muslims tend to lack democratic institutions but not democratic inclinations (Bratton 2003; Hoffman 2004; Inglehart and Norris 2003; Jamal and Tessler 2008; Mogahed 2006; Rose 2002; Rowley and Smith 2009). Rather to the contrary, analysis of SP data from the World Values Survey (WVS) reveals that both individual Muslims and Muslim-majority countries have relatively positive attitudes towards democracy (Inglehart and Norris 2003; Hoffman 2004; Rowley and Smith 2009).

To show this, we repeat Rowley and Smith's (2009: 290-292) individual-level analysis of the association between Islam and democratic attitudes—which is the most recent one—at the country level. Data on attitudes towards democracy also come from the WVS. To ensure direct comparability with the Rowley and Smith (2009) results we do not use data from the most recent WVS wave (World Values Association 2009) using the first four waves instead (European Values Study Group and World Values Survey Association 2006). Typically, we have about 145,000 individuals residing in some 80 countries. Following Rowley and Smith's (2009: 285, 287) classification there are 13 Muslim countries in our sample. According to data from the CIA World Factbook (Central Intelligence Agency 2009), these countries all have sizable Muslim majorities. Some countries have been sampled more than once so that the total number of country-level observations is more than 100. Table 3.1 depicts the results for the five items on democracy included in the WVS.[29]

[29] Rowley and Smith's (2009: 290-292) individual-level evidence based on the same items concerning individuals' attitude towards democracy can be found in their Tables 7 to 11. Their Table 12 (Rowley and Smith 2009: 293) gives some aggregate-level evidence of the positive relation between Islam and attitudes towards democracy.

Table 3.1: Support for Democracy in Muslim and Non-Muslim Countries.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them? In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	81	79	79	80	79
Muslim countries	Average	3.50 (.169)	2.86 (.125)	2.64 (.286)	2.80 (.164)	3.34 (.293)
	4	59.4% (12.2%)	18.9% (7.2%)	15.2% (6.3%)	18.5% (6.4%)	48.9% (18.3%)
	3	33.4% (12.3%)	55.9% (14.0%)	45.0% (15.5%)	51.4% (12.2%)	39.8% (13.3%)
	2	5.5% (7.5%)	17.9% (6.9%)	28.7% (12.9%)	21.5% (7.5%)	7.7% (6.2%)
	1	1.8% (1.6%)	7.3% (5.1%)	11.1% (8.8%)	8.5% (6.1%)	3.7% (4.0%)
	Observations	16	16	16	16	16
Non-Muslim countries	Average	3.32 (.222)	2.71 (.241)	2.47 (.234)	2.72 (.245)	3.23 (.226)
	4	45.0% (15.9%)	13.8% (8.8%)	9.7% (5.9%)	15.1% (9.1%)	38.1% (15.6%)
	3	44.3% (12.1%)	51.5% (12.5%)	38.6% (11.4%)	49.2% (11.9%)	49.0% (12.5%)
	2	8.0% (5.0%)	26.9% (9.3%)	40.1% (9.2%)	27.9% (10.4%)	10.4% (6.8%)
	1	2.7% (1.9%)	7.7% (6.7%)	11.5% (8.5%)	7.8% (5.9%)	2.5% (2.4%)
	Observations	105	102	102	103	102

Notes: Table reports aggregated individual answers for different countries and years. Muslim countries included are Albania, Algeria, Azerbaijan, Bangladesh, Indonesia, Iran, Iraq, Jordan, Morocco, Pakistan, Saudi Arabia (not the first item), Turkey and Egypt (cf. Rowley and Smith 2009: 285, 287). Answers range from 1 (least democratic response) to 4 (most democratic response). Answer categories for the last item (Item 5) have been recoded accordingly. Standard deviations in parentheses. Due to rounding, percentages may not add up to 100%. See Section 3.4 and Table 3.2 for further details about the items.

It appears that democracy has universal appeal with the percentages of people viewing democracy as something very good or fairly good being high both among Muslim and non-Muslim societies; the third item asking people whether they feel democracy is indecisive and has too much squabbling is an exception. It seems the difference between Muslim countries and the rest of the world does not lie so much in the share of democratically inclined people but in how positive towards democracy people are. Whereas almost everyone seems to view democracy favorably to some extent, Table 3.1 clearly shows that people from Muslim countries are more prone to be very rather than fairly positive about democracy. The most comprehensive evidence comes from comparing the mean scores on the five democracy items. In most cases, the fourth item is an exception, the average attitude towards democracy is statistically significantly more positive in Muslim countries than it is in other countries ($p < 0.10$ or better). It are these latter two findings that make the relation between Islam and democracy appear paradoxical; "Muslims claim to like democracy, so why do they have so little?" (Rowley and Smith 2009: 273).

This (alleged) paradox is robust to changes in the criterion that identifies countries as Muslim. In the appendix we present results for classifications based on Islam as the largest religion, Islam as religion of over 90% of the population, or predominantly Sunni Muslim populations, amongst others. We find that replacing Rowley and Smith's (2009: 285, 287) country classification with these criteria does not qualitatively affect the results and hence does not account for the observed paradox.

Since they are tapping political attitudes in states where people may feel they are being endangered by expressing their political views, we obviously need to take results of the surveys in question with a healthy dose of caution (Kuran 1995; cf. Rowley and Smith 2009: 274; see, also, Bertrand and Mullainathan 2001, and our brief discussion of well-known biases in SP measures in the previous chapter). However, whilst a repressive environment may lead to preference falsification and thus cast doubt on the validity of the SP method or other data obtained using questionnaires, it is not clear that this should have a systematic effect and thereby translate into a significantly more positive evaluation of democracy. In some cases, particularly where the authoritarian regime makes a point of cultivating a democratic façade, it might. It is just as likely, however, that repression causes people to understate their commitment to democracy. Hence, biases due to repressive environments likely do not cause the positive relation between Muslim identity and democratic attitudes. We need a different explanation.

3.4. EXPLAINING AWAY THE DEMOCRACY PARADOX

3.4.1. The Diminishing Marginal Utility of Democracy

Can the principle of diminishing marginal utility, which Inglehart (1997) argues is the driving force behind the famous value shift in industrial societies, explain the other paradoxes in the study of (heterogeneous) stated preferences and differences in sociopolitical and socioeconomic outcomes? The answer seems to be yes. The fact that respondents in Muslim-majority countries tend to favor democracy more, while having less of it, suddenly makes sense if one interprets the questionnaire items as eliciting a marginal preference. If values survey scores depict marginal preferences rather than an underlying cultural attitude, Muslims' sympathy for democracy simply reflects their relative deprivation of it; other people would be similarly enthusiastic about democracy were they to live in the kind of repressive environments that characterize many Islamic countries.

From this perspective, the paradox observed in studies like those of Hoffmann (2004), Inglehart and Norris (2003), and Rowley and Smith (2009) needs no explanation. It simply does not exist. Countries with relatively low levels of democracy are likely to favor democracy more strongly because people value that which is scarce. To see a paradox in this is a consequence of a misinterpration of survey items as eliciting cultural attitudes whereas in actuality they capture marginal preferences.

In the remainder of this section we formulate some hypotheses applying the marginal preferences logic to account for the alleged Muslim democracy paradox and test these empirically using the same WVS data that we used for Table 3.1.

3.4.2. Hypotheses

It is possible to test whether the marginal preferences problem plaguing SP data accounts for the paradoxical relation between Islam and democracy by looking at the relation between the WVS items measuring democratic attitudes and democratic institutions in general. If the WVS items elicit cultural attitudes, as they have been designed to do, we expect a positive relation between sympathy for democracy and democratic institutions. In that case, the paradox remains; the question is why Muslim-majority countries' institutions do not evolve in line with their populations' preferences. If, on the other hand, the WVS items elicit marginal preferences, we expect a negative correlation between the level of democracy and democratic

attitudes as measured by the WVS items. The finding that Muslim-majority countries combine a relatively strong preference for democracy with little actual democracy in that case confirms the negative relation between marginal preferences and circumstances, which, in turn, is indicative of the marginal preferences problem plaguing the SP method; the more one has of something, the less one desires increments to it.

At the same time, prosperity and education likely foster a more favorable outlook on democracy. This holds independent of the question whether survey measures elicit cultural attitudes or marginal preferences, although the mechanism differs between the two constructs. In the case of cultural attitudes, we expect the effect of gross domestic product (GDP) and education to run primarily through factors such as literacy and the size of a critical middle class, changing the mindset of people. In the case of marginal preferences, we also expect an effect of GDP and education, but here the mechanism is that satiation of these rival objectives makes democracy more important to people by default, rather like satiating materialist objectives makes people more postmaterialist (Inglehart 1987, 1990, 1997). In line with Barro's view of democracy as a sort of luxury good (Barro 1996: 24), the relative deprivation of democracy is larger in societies where other pressing needs have been met.

Regardless of the mechanism involved, we therefore expect a positive effect of income and education on attitudes towards democracy. Any relation between democratic attitudes and levels of democracy should subsequently be seen as conditional on education and individual and country per-capita income. We have the following two competing hypotheses:

Hypothesis 1a *If the World Values Survey items used to measure democratic attitudes elicit cultural attitudes, we expect a positive relation between attitudes measured by these items and actual levels of democracy, other factors controlled for.*

Hypothesis 1b *If the World Values Survey items used to measure democratic attitudes elicit marginal preferences, we expect a negative relation between attitudes measured by these items and actual levels of democracy, other factors controlled for.*

We test these hypotheses below where we combine individual-level data on professed attitudes towards democracy with data on existing levels of democracy in respondents' countries, and include variables concerning individuals' sociodemographic characteristics and

other features of their societies as controls, particularly individuals' level of education and GDP per capita.

3.4.3. Empirical Analysis

Data

Data on our dependent variables come from the World Values Survey or WVS. The WVS has grown out of the European Values Survey (EVS) and is currently coordinated by Ronald Inglehart (e.g. Inglehart 1997). During the period 1981-1984 the EVS surveyed respondents from a number of European countries, but since then the WVS has evolved into a global project covering almost 100 societies. Together these span some 90% of world population. The websites of the WVS and the EVS provide more information.[30] We follow existing studies arguing for a Muslim democracy paradox, specifically Rowley and Smith (2009), and use responses from the first four waves of the survey (see above). Interviews for Wave 1 took place between 1981 and 1984, and covered 20 countries and roughly 25,000 individuals. Wave 2 was conducted between 1989 and 1993, and included 42 countries for a total of some 61,000 respondents. Wave 3 followed immediately thereafter, covering the period 1994 to 1998 and including 52 countries and about 75,000 respondents in its sample. Finally, Wave 4 spans the period 1999 to 2004, during which approximately 96,000 respondents from 67 countries were interviewed. Data files (European Values Study Group and World Values Survey Association 2006; World Values Survey Association 2009) are publicly available from the websites of the organizations involved.

For our research, it is particularly important to note that in the design and conduct of the WVS care is taken to minimize the risk of potential distortions caused by repressive environments (see our discussion on preference falsification in Section 3.2 and, again, the more general discussion of social desirability and other biases in SP data in the fourth section of Chapter 2). Survey questions are largely identical between countries. No compromises were made in politically repressive regimes and interviewers are instructed to stress the confidentiality of respondents' answers. Interviews are conducted in isolation. In strongly male-dominated societies, women are interviewed by female interviewers. The websites of the WVS and the EVS again give further information concerning the items included in the survey and the way country surveys are conducted.

[30] See http://www.worldvaluessurvey.org and http://www.europeanvaluesstudy.eu.

The WVS dataset contains several items concerning individuals' attitudes towards democracy. Following Rowley and Smith (2009) we identify five of these (cf. Table 3.1). The first item, Item 1, asks individuals how they feel about different systems for governing their country, including democracy:

> *I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? [Having a democratic political system]*

The other four items ask respondents about their attitude towards certain features of democracy as a political system:

> *I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them?*

The specific items read as follows:

- *In democracy, the economic system runs badly* (Item 2);
- *Democracies are indecisive and have too much quibbling* (Item 3);
- *Democracies aren't good at maintaining order* (Item 4); and
- *Democracy may have problems but it's better than any other form of government* (Item 5).

We recode answers to these variables, specifically for the last item, so that a score of 1 corresponds to the least democratic response and 4 to the most democratic response with 2 and 3 in between (again see Table 3.1).

Next to data on democratic attitudes, we further rely on the WVS dataset for individual-level control variables. Our aim is to investigate whether a country's level of democracy has a negative effect on the value its citizens attach to democracy and we do so with potential individual determinants of democratic attitude controlled for. These are employment status (Full-time (=base category), Part-time, Self-employed, Retired, Housewife, Student, Unemployed, or Other), income scale (1-10), education (Inadequately

completed elementary education, Completed (compulsory) elementary education (=base category), Incomplete secondary school: technical/vocational type, Complete secondary school: technical/vocational type, Incomplete secondary: university-preparatory type, Complete secondary: university-preparatory type, Some university without degree, or University with degree), sex, and age and age squared. Table 3.2a gives descriptive statistics.

Table 3.2a: Descriptive Statistics.

Variable and description	Mean and standard deviation
Dependent variables	
Item 1 [155,879]	3.34 (.733)
Item 2 [141,203]	2.73 (.803)
Item 3 [143,891]	2.48 (.833)
Item 4 [144,305]	2.72 (.820)
Item 5 [146,381]	3.24 (.743)
Level-2 independent variables (between-countries)	
Polity IV [152,364]	5.66 (5.50)
GDP per capita in 1990$ [165,058]	8,541 (7,054)
Level-1 independent variables (within-countries)	
Sex (male = 1) [167,072]	48.9% (50.0%)
Age [166,849]	40.8 (16.0)
Employment status [162,769]	
Full-time employed	36.4% (48.1%)
Part-time employed	7.5% (26.3%)
Self-employed	9.5% (29.3%)
Retired	13.8% (34.4%)
Housewife	13.9% (34.6%)
Student	7.5% (26.4%)
Unemployed	9.5% (29.3%)
Other employment status	2.0% (14.0%)
Income scale (1-10) [146,204]	4.59 (2.47)
Education [163,848]	
Inadequately completed elementary education	10.6% (30.8%)
Completed (compulsory) elementary education	15.1% (35.8%)
Incomplete secondary: technical/vocational	9.1% (28.8%)
Complete secondary: technical/vocational	17.2% (37.7%)
Incomplete secondary: university-preparatory	9.7% (29.6%)
Complete secondary: university-preparatory	16.6% (37.2%)
Some university without degree	7.5% (26.4%)
University with degree	14.2% (34.9%)

Notes: Standard deviations in parentheses and number of observations in square brackets. Respondents with "don't know" or otherwise missing answers have been dropped. Data on selected independent variables cover only individuals and country observations for which scores on the dependents, Item 1-5, are available.

We supplement the WVS data on attitudes towards democracy and sociodemographics with country (and year) data on democracy and per-capita GDP. The source of the democracy

data is the Polity IV Project (Marshall 2008). This project measures the democratic quality of governing institutions for 163 countries over the period 1800-2008. We use its revised polity measure (the so-called polity 2 measure) of democracy. This is our key independent variable. We match the polity measure of democracy and our individual measures of democratic attitudes both to countries and years of the WVS data. Scores on the polity measure of democracy from the Polity IV Project (Marshall 2008) range from -10 (lowest score possible) to +10 (highest score possible). The Polity IV website gives a detailed description of the polity measure of democracy and access to a downloadable data file.[31]

Table 3.2b: Pairwise Correlations Between Selected Variables.

Variable	Polity	GDP [1990$]
Polity	1	
GDP [1990$]	.469	1
Inadequately completed elementary education	-.075	-.122
Completed (compulsory) elementary education	.060	.043
Incomplete secondary: technical/vocational	.082	.053
Complete secondary: technical/vocational	-.062	-.016
Incomplete secondary: university-preparatory	.032	.025
Complete secondary: university-preparatory	-.034	-.036
Some university without degree	.037	.074
University with degree	-.017	-.003

Notes: See Table 3.2a.

Our GDP data come from The Conference Board & Groningen Growth and Development Center (2008) and we again make an exact match between countries and years available from this dataset and our WVS data. Levels of GDP per capita range from about $600 to almost $34,000 (1990 PPPs). The website of the Conference Board has further information and a data file available for download.[32] The bottom two rows of Table 3.2a give descriptive statistics for these two country-level variables. Table 3.2b shows the correlations between these and some selected individual-level variables. Correlations are low to moderate, except between democracy and GDP per head, which is as expected. We check the robustness of our results to a possible multicollinearity problem due to the high correlation between the polity measure and national income.

[31] See http://www.systemicpeace.org/polity/polity4.htm.

[32] See http://www.conference-board.org/economics/database.cfm.

Method

Our preferences data concern the attitudes towards democracy expressed by individuals living in certain societies and faced with a certain political environment. The nature of the data, and our theoretical argument predicting that contextual factors shape individuals' scores on questionnaire items aiming to elicit attitudes, call for a multilevel approach to testing our hypotheses. Using multilevel or hierarchical linear modeling (Gelman and Hill 2007; Raudenbush and Bryk 2002; Snijders and Bosker 1999) we avoid simply throwing together predictors at the individual and the contextual level, and treating them the same. In our case, we have individuals (Level 1) who are nested in countries (Level 2), and applying a multilevel technique we can separate the individual and aggregate correlates of democratic attitudes and model them simultaneously. Multilevel modeling also allows for more efficient inference than is possible with either complete pooling or no pooling of the data. Given the nature of the data, with individuals nested in countries, it is further likely that clustering of observations causes a problem for traditional regression techniques. Individuals within countries are not independent observations so standard errors may be underestimated (Moulton 1986, 1990; Wooldridge 2003). Multilevel modeling takes such clustering into account.

For the formal empirical model we start with an individual i (Level 1) living in country j (Level 2). DA_{ij} subsequently denotes the individual's attitude towards democracy (any of the five items). The independent variable of interest is the level of democracy in a country measured by its score on the polity measure of democracy, PD_j. We also include x_{ij}, a set of individual-level control variables, and GDP_j, a country's level of per-capita GDP as a country-level control variable. This gives the following empirical models for Levels 1 (within countries) and 2 (between countries), separately and combined:

$$
\begin{aligned}
&\text{Level 1:} \\
&\quad DA_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \\
&\text{Level 2:} \\
&\quad \beta_{0j} = \gamma_{00} + \gamma_{01} PD_j + \gamma_{02} GDP_j + u_{0j} \\
&\text{Complete:} \\
&\quad DA_{ij} = \gamma_{00} + \gamma_{01} PD_j + \gamma_{02} GDP_j + \beta_1 x_{ij} + \left(u_{0j} + \varepsilon_{ij}\right)
\end{aligned}
\qquad . \qquad (3.1)
$$

The complete general model is a varying-intercepts model. Any country-specific factor that may affect individuals' attitude towards democracy but is not controlled for by the aggregate-level variables that we include is contained in these random intercepts. The terms in parentheses constitute the random part of the model and the other terms the fixed part. The error terms in the random part capture the difference between a multilevel model and a traditional model: there is a "normal" residual error term (ε_{ij}), familiar from classic regression analysis, but also an aggregate-level error term (u_{0j}). The model is estimated using maximum likelihood procedures.

Results

Table 3.3 gives the results for our baseline model with only individual-level (control) variables included. We estimate five empirical models (A1-A5) and include all of the individual variables. Hence, the models differ only with respect to their dependent variable, i.e., the specific survey item used to measure democratic attitude. In line with the existing literature and our predictions, both personal income and education contribute to a more positive attitude towards democracy. In addition, we find that being employed generally has a positive effect, and that men tend to have stronger democratic attitudes than women do.

Table 3.3: Baseline Model with Individual Determinants of Democratic Attitude.

Variables	Model A1	Model A2	Model A3	Model A4	Model A5
Intercept	3.10***	2.52***	2.31***	2.54***	2.97***
	(.033)	(.034)	(.037)	(.035)	(.036)
Part-time	-.012	-.024**	-.002	-.001	.002
	(.008)	(.009)	(.010)	(.010)	(.009)
Self-employed	.008	.004	-(.008	-.008	.013
	(.008)	(.009)	(.009)	(.009)	(.008)
Retired	-.024***	-.052***	-.067***	-.045***	-.027***
	(.009)	(.010)	(.010)	(.010)	(.009)
Housewife	.002	.001	.015	-.006	.006
	(.008)	(.009)	(.009)	(.009)	(.008)
Student	.049***	.018	.029**	.025**	.021**
	(.009)	(.011)	(.011)	(.011)	(.010)
Unemployed	-.026***	-.057***	-.042***	-.037***	-.007
	(.008)	(.009)	(.009)	(.009)	(.008)
Other employment status	-.014	-.016	-.013	.005	-.003
	(.015)	(.017)	(.017)	(.017)	(.015)
Income scale (1-10)	.010***	.020***	.017***	.016***	.008***
	(.001)	(.001)	(.001)	(.001)	(.001)
Inadequately completed elementary education	-.003	-.007	.035***	.015	.003
	(.009)	(.010)	(.011)	(.011)	(.009)

Table 3.3, continued.

Variables	Model A1	Model A2	Model A3	Model A4	Model A5
Incomplete secondary: technical/vocational	-.005 (.009)	-.023** (.010)	-.019* (.010)	-.001 (.010)	.005 (.009)
Complete secondary: technical/vocational	.070*** (.008)	.053*** (.009)	.034*** (.009)	.063*** (.009)	.054*** (.008)
Incomplete secondary: university-preparatory	.068*** (.009)	.085*** (.010)	.064*** (.010)	.090*** (.010)	.083*** (.009)
Complete secondary: university-preparatory	.125*** (.008)	.103*** (.009)	.095*** (.009)	.121*** (.009)	.105*** (.008)
Some university without degree	.188*** (.010)	.188*** (.011)	.153*** (.011)	.208*** (.011)	.176*** (.010)
University with degree	.233*** (.008)	.244*** (.009)	.219*** (.010)	.250*** (.010)	.206*** (.009)
Male	.048*** (.004)	.060*** (.005)	.019*** (.005)	.033*** (.005)	.027*** (.005)
Age [/100]	.251*** (.079)	.018 (.091)	.150 (.093)	.123 (.093)	.378*** (.083)
Age^2 [/10,000]	-.048 (.089)	.024 (.102)	-.207** (.104)	-.240** (.104)	-.122 (.093)
Countries	69	68	68	68	68
Level-2 observations	101	98	98	98	98
Individual observations	117,565	106,031	107,824	107,240	109,335
-2Loglikelihood	247,442.9	242,503.9	254,248.1	249,992.7	234,004.7

Notes: See Table 3.2. Standard errors in parentheses. * (**) (***) indicates significance at the 10% (5%) (1%) level. Base category is a full-time employed woman with completed (compulsory) elementary education. All models have varying (i.e. random) intercepts.

To test our hypotheses, we estimate the same five models, further including GDP per capita and the actual level of democracy as determinants of democratic attitudes (Models B1-B5). As Table 3.4 shows, there is a clear negative relation between the level of democracy and the attitude towards democracy for four out of five measures. For the other measure, Item 1, we find no statistically significant relation between attitudes and levels of democracy. In line with our predictions, GDP per capita has a positive effect.

The significance of GDP and the polity measure suggests that B models provide improved statistical fits over A models. This is confirmed by Likelihood ratio tests, which show that all B models provide a statistically significantly better fit than the A models ($p < 0.05$ or better). To be exact, the B models are nested in the A models so that the test-statistic is obtained by simply detracting the -2Loglikelihood of the nested model from that of the original model, e.g. 247,442.9 - 247,435.8 = 7.1 (Model A1 vs. Model B1). The probability distribution of this test statistic is approximately chi-square with, in our case, two degrees of freedom (we added the polity measure and GDP as explanatory variables).

The findings reported in Tables 3.3 and 3.4 lead us to accept Hypothesis 1b (and reject Hypothesis 1a). Democratic attitudes do not seem to determine actual levels of democracy. Rather, they appear to respond to the level of democracy; if there is enough, people value it less. We already noted, however, that the correlation between GDP per capita and level of democracy may result in a multicollinearity problem. Hence, to check the robustness of our findings we have also estimated the effect of the actual level of democracy on professed attitude towards democracy without controlling for levels of per-capita income.

Table 3.4: Multilevel Results of the Relation between Level of Democracy and Democratic Attitude.

Variables	Model B1	Model B2	Model B3	Model B4	Model B5
Intercept	3.03***	2.44***	2.20***	2.42***	3.16***
	(.043)	(.041)	(.050)	(.043)	(.065)
Polity	-.001	-.022***	-.022***	-.013***	-.059***
	(.001)	(.002)	(.002)	(.002)	(.002)
GDP	.078***	.217***	.249***	.205***	.146***
	(.029)	(.029)	(.035)	(.030)	(.039)
Individual controls	Yes	Yes	Yes	Yes	Yes
-2Loglikelihood	247,435.8	242,352.7	254,112.2	249,925.1	233,032.9

Notes: See Table 3.3. Individual controls are employment status, income scale (1-10), sex, education (dummies), age, and age^2.

The model without GDP is not our preferred model for theoretical reasons outlined above. Nevertheless, Table 3.A.6 (Models C1-C5) in the appendix shows our results to be robust: also with GDP excluded does the actual level of democracy have a negative effect on attitudes towards democracy. This holds for Models C2-C5; the first democracy item again is an exception. As we would expect, Models C1-C5 provide better fits than Models A1-A5, but poorer fits than Models B1-B5, as evidenced by their -2Loglikelihood score. Likelihood ratio tests show the difference to be statistically significant ($p < 0.01$). Model B1 is the exception.

Turning to implications for the democracy paradox, we find that no such paradox exists. Firstly, the paradox holds much more broadly. Countries that have lower levels of democracy are consistently more democratically inclined—this negative correlation is not limited to Muslim countries. Secondly, measurement issues likely account for this observed negative correlation between attitudes towards democracy and actual democratic practice: such a negative attitude-practice correlation may result from the marginal-preferences problem—in which respondents are led by their current circumstances when asked how they feel about a certain state of affairs—known to affect attitudinal measures.

3.5 CONCLUSION AND DISCUSSION

Is there a democracy paradox of Islam? This paper finds that the empirical support for such a paradox does not stand up to critical scrutiny. Our contribution goes further, however, as we present convincing evidence that the SP method is prone to mismeasure heterogeneity for instance in cultural attitudes. Concerning the democracy paradox of Islam, our analysis demonstrates that the findings that Islamic countries are less democratic, whereas their inhabitants value democracy more, are not exceptional. A negative relation between democratic attitudes and democracy levels is a general phenomenon, not specific to Muslim countries. Neither is such a negative relation paradoxical. A relative lack of democracy makes people want more of it. This is the principle of diminishing marginal utility. That principle applies to marginal preferences, however, and not to people's underlying attitudes. It is only the preference for *more* democracy that increases when democratic institutions fade; the attitude towards democracy *in general* is not affected.

The results of our empirical scrutiny of the supposed democracy paradox support the idea that the SP method in general, and values surveys in particular, suffer a marginal preferences problem: questionnaire items on attitudes and related entities tend to elicit marginal preferences (the importance attached to an objective on top of its current level) rather much more so than durable attitudinal traits (the importance attached to an objective in general). This is not to say that religious and cultural values do not matter for democracy or other sociopolitical outcomes. It does imply, however, that if we want to investigate cross-cultural differences in attitudes towards democracy or any other objective, the SP method and values survey items have limited usefulness.

The above conclusion generalizes to the use of SP measures in the study of the economics of preference heterogeneity at large. To see this, let us briefly discuss the findings of this chapter using terms and ideas introduced in the fourth section of the previous chapter. The first thing we then conclude is that SP measures of preference heterogeneity simply lack psychometric quality, specifically (measurement/construct) validity. SP measures by and large do not elicit value preferences or attitudes, capturing marginal preferences, which are a different construct altogether, instead.[33] The second thing we conclude is that SP measures have limited explanatory capacity. The marginal preferences problem means that ascribing

[33] This chapter has not addressed issues of reliability, but see Clarke et al. (1999) and Duch and Taylor (1993), amongst others.

differences in whichever attitudes, to religion may be misleading. The reason is that differences in professed attitudes may not depend on religious or cultural differences between groups but likely have a rationale in the different circumstances in which these groups find themselves. This, in turn, poses a problem for scientific explanation, akin to the explanans-explanandum mix up that we discussed at length in Chapter 2. Simply put, if attitudinal survey scores are indeed determined by sociopolitical and socioeconomic context, logically they cannot simultaneously explain differences in such circumstances. To do so would result in a circular argument or, worse, in conclusions that are plain wrong.

We can summarize the above in the form of an update to Table 2.1. For this purpose, Table 3.5 repeats our evaluation of the usefulness of the SP approach and SP measures in the study of the economics of preference heterogeneity from the previous chapter and adds our new findings. We have not presented novel evidence concerning SP measures' availability, scope, and resolution, all aspects of such measures' usability. Following the above discussion, we have lowered scores on psychometric quality (reliability and validity) and explanatory capacity, the latter also an aspect of usability. Hereby it is interesting to note that the low score on explanatory capacity is due to the low score on psychometric quality and, specifically, to the source of the latter low score, namely a substantial lack of construct validity. In the previous chapter we emphasized the potential for trade-offs between psychometric quality and usability, but apparently such trade-offs are not a given. Psychometric quality can be an important prerequisite for the usefulness of measures of heterogeneous preferences and probably other measures of psychological constructs as well.

Table 3.5: A Summary Evaluation of the Usefulness of the SP Approach to Measuring Heterogeneous Preferences and Studying the Economics of Preference Heterogeneity.

	New evaluation scores	Old evaluation scores
Psychometric quality	●	● - ●●●
Usability		
Availability	●●●●●	●●●●●
Scope	●●●●●	●●●●●
Resolution	●●●●●	●●●●●
Explanatory capacity	●	● - ●●●●

Notes: See Table 2.1.

Three final remarks are meant to nuance our conclusions a bit. Firstly, our finding that SP data suffer the marginal preferences problem does not imply that all SP measures are equally unsuccessful at eliciting values. Typically, we expect SP items to elicit values as well as marginal preferences, both to some unknown extent (cf. Maseland and van Hoorn 2009).

Very often the marginal preference part will dominate, resulting in the by now familiar negative values-practices correlations. The results from the GLOBE project (House et al. 2004: 736)—negative correlations in seven out of its nine dimensions of culture—suggest that this will be the case about 75% of the time.[34] If values dominate, correlations are bound to be positive, whereas absent correlations suggest that neither construct dominates. Overall, given the likelihood of SP items capturing both values and marginal preferences, most of the time it will not be clear what exactly one is measuring. Secondly, concerning SP measures' limited usefulness we do not find that SP measures, specifically when capturing marginal preferences, have no usefulness whatsoever (cf. Maseland and van Hoorn 2010a)—our conclusion holds in the context of using SP measures in the study of the economics of preference heterogeneity, but not necessarily outside these boundaries. Finally, by arguing that marginal preferences lie behind the observed democracy paradox of Islam we do not mean to claim that this is the only possible answer to this puzzle. As pointed out, survey instruments need to be treated with caution when used in politically repressive contexts. We therefore conclude with a plea for further research in this direction.

[34] To be exact, the dimensions showing a negative correlation are "Assertiveness" (r = -0.26), "Institutional collectivism" (r = -0.61), "Future orientation" (r = -0.41), "Humane orientation" (r = -0.32), "Performance orientation" (r = -0.28), "Power distance" (r = -0.43), and "Uncertainty avoidance" (r = -0.62). All these correlations were significant at the 5% level (n = 61). The only dimension showing a statistically significantly positive correlation was "Gender egalitarianism" (r = 0.32). Finally, "In-group collectivism" shows a positive but statistically insignificant correlation (r = 0.21). Detailed descriptions of these dimensions can be found in several of the project's published works (e.g. House et al. 2004; Chokar et al. 2007).

Table 3.A.1: Support for Democracy in Countries with More than 90% Muslim Population.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them?			
			In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	81	79	79	80	79
Muslim countries	Average	3.43 (.145)	2.79 (.092)	2.49 (.274)	2.71 (.145)	3.26 (.239)
	4	56.7% (10.2%)	21.2% (6.9%)	15.1% (7.4%)	18.8% (7.6%)	46.3% (13.3%)
	3	32.1% (11.7%)	46.6% (9.3%)	34.0% (9.5%)	44.3% (9.5%)	3.8% (9.1%)
	2	8.6% (9.0%)	22.5% (3.4%)	36.2% (11.0%)	25.9% (5.8%)	9.5% (4.8%)
	1	2.5% (1.7%)	9.8% (4.6%)	14.8% (7.3%)	11.0% (5.4%)	5.4% (4.2%)
	Observations	9	9	9	9	9
Non-Muslim countries	Average	3.33 (.229)	2.73 (.241)	2.49 (.247)	2.73 (.244)	3.24 (.238)
	4	46.1% (16.4%)	14.0% (8.6%)	10.1% (6.0%)	15.3% (8.9%)	39.0% (16.5%)
	3	43.8% (12.3%)	52.6% (12.9%)	39.9% (12.2%)	50.0% (12.1%)	48.5% (13.0%)
	2	7.6% (5.1%)	26.0% (9.8%)	38.8% (10.5%)	27.1% (10.6%)	10.1% (6.9%)
	1	2.6% (1.9%)	7.5% (6.6%)	11.2% (8.6%)	7.6% (5.9%)	2.4% (2.4%)
	Observations	112	109	109	110	109

Notes: See Table 3.1. Countries with more than 90% Muslim population included in the WVS dataset are Algeria, Iran, Iraq, Jordan, Morocco, Pakistan, and Turkey. Data on Muslim share in total population come from the CIA World Factbook (Central Intelligence Agency 2009).

Table 3.A.2: Support for Democracy in Countries in which Islam is the Largest Religion.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them? In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	81	79	79	80	79
Muslim countries	Average	3.49 (.186)	2.84 (.153)	2.62 (.263)	2.79 (.207)	3.29 (.324)
	4	59.1% (14.0%)	19.5% (8.0%)	15.1% (6.4%)	19.6% (8.5%)	46.7% (18.9%)
	3	33.3% (12.7%)	53.3% (13.3%)	42.7% (14.1%)	49.0% (12.2%)	39.5% (13.4%)
	2	5.6% (6.7%)	19.2% (7.3%)	30.8% (12.0%)	22.4% (9.1%)	9.5% (8.9%)
	1	2.0% (1.5%)	7.9% (4.8%)	11.4% (7.9%)	9.0% (5.8%)	4.2% (4.1%)
	Observations	22	22	22	22	22
Non-Muslim countries	Average	3.31 (.219)	2.71 (.242)	2.46 (.237)	2.71 (.242)	3.23 (.214)
	4	44.2% (15.4%)	13.4% (8.5%)	9.4% (5.7%)	14.7% (8.7%)	37.9% (15.3%)
	3	45.0% (11.6%)	51.9% (12.6%)	38.7% (11.6%)	49.7% (11.9%)	49.7% (12.2%)
	2	8.1% (5.0%)	27.2% (19.3%)	40.4% (9.3%)	28.1% (10.3%)	10.1% (6.2%)
	1	2.7% (1.9%)	7.6% (6.9%)	11.5% (8.7%)	7.6% (5.9%)	2.3% (2.1%)
	Observations	99	96	96	97	96

Notes: See Table 3.A.1. Countries with Islam as the dominant religion included in the WVS dataset are Albania, Algeria, Azerbaijan, Bangladesh, Bosnia and Herzegovina, Indonesia, Iran, Iraq, Jordan, Kyrgyzstan, Morocco, Nigeria, Pakistan, Saudi Arabia (not the first item), Turkey, Egypt, and Tanzania.

Table 3.A.3: Support for Democracy in Arab and Non-Arab Countries.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them? In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	81	79	79	80	79
Arab countries	Average	3.57 (.137)	2.83 (.127)	2.47 (.286)	2.77 (.154)	3.37 (.261)
	4	63.9% (11.5%)	24.6% (6.1%)	15.3% (6.4%)	23.3% (5.4%)	52.2% (16.0%)
	3	30.7% (10.4%)	44.6% (8.9%)	32.7% (11.4%)	42.1% (9.4%)	36.3% (10.7%)
	2	3.5% (1.7%)	20.0% (3.5%)	35.9% (8.3%)	22.4% (3.1%)	7.5% (5.4%)
	1	1.9% (1.4%)	10.7% (5.1%)	16.1% (9.2%)	12.2% (6.1%)	4.0% (3.1%)
	Observations	5	6	6	6	6
Non-Arab countries	Average	3.33 (.223)	2.73 (.273)	2.49 (.247)	2.73 (.241)	3.24 (.235)
	4	46.1% (16.0%)	14.0% (8.5%)	10.2% (6.1%)	15.2% (8.8%)	38.9% (16.1%)
	3	43.4% (12.5%)	52.5% (12.8%)	39.8% (12.1%)	49.9% (12.0%)	48.4% (12.8%)
	2	7.8% (5.4%)	26.0% (9.6%)	38.7% (10.6%)	27.3% (10.5%)	10.2% (6.8%)
	1	2.6% (1.9%)	7.5% (6.5%)	11.2% (8.5%)	7.6% (5.8%)	2.5% (2.7%)
	Observations	116	112	112	113	112

Notes: See Table 3.A.1. Arab countries included in the WVS dataset are Algeria, Iraq, Jordan, Morocco, Saudi Arabia (not the first item), and Egypt.

Table 3.A.4: Support for Democracy in Arab and Non-Arab Muslim Countries.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them? In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	12	13	13	13	13
Arabic Muslim countries	Average	3.57 (.137)	2.83 (.127)	2.47 (.286)	2.77 (.154)	3.37 (.261)
	4	63.9% (11.5%)	24.6% (6.1%)	15.3% (6.4%)	23.3% (5.4%)	52.2% (16.0%)
	3	30.7% (10.4%)	44.6% (8.9%)	32.7% (11.4%)	42.1% (9.4%)	36.3% (10.7%)
	2	3.5% (1.7%)	20.0% (3.5%)	35.9% (8.3%)	22.4% (3.1%)	7.5% (5.4%)
	1	1.9% (1.4%)	10.7% (5.1%)	16.1% (9.2%)	12.2% (6.1%)	4.0% (3.1%)
	Observations	5	6	6	6	6
Non-Arabic Muslim countries	Average	3.48 (.181)	2.88 (.126)	2.75 (.242)	2.82 (.174)	3.32 (.323)
	4	57.4% (12.5%)	15.4% (5.5%)	15.2% (6.5%)	15.7% (5.2%)	46.8% (20.1%)
	3	34.6% (13.3%)	62.7% (12.2%)	52.3% (13.0%)	57.1% (10.2%)	41.8% (14.8%)
	2	6.3% (9.0%)	16.6% (8.2%)	24.4% (13.6%)	21.0% (9.3%)	7.8% (6.9%)
	1	1.7% (1.7%)	5.2% (4.1%)	8.0% (7.4%)	6.3% (5.2%)	3.5% (4.5%)
	Observations	11	10	10	10	10

Notes: See Table 3.A.1. Arab Muslim countries included in the WVS dataset are Algeria, Iraq, Jordan, Morocco, Saudi Arabia (not the first item), and Egypt.

Table 3.A.5: Support for Democracy in Muslim and Non-Muslim Countries with Shia Countries Excluded.

		I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good (4), fairly good (3), fairly bad (2) or very bad (1) way of governing this country? "Having a democratic political system"	I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly (1), agree (2), disagree (3) or disagree strongly (4), after I read each one of them? In democracy, the economic system runs badly	Democracies are indecisive and have too much quibbling	Democracies aren't good at maintaining order	Democracy may have problems but it's better than any other form of government
	Countries	79	77	77	78	77
Muslim countries	Average	3.52 (.174)	2.86 (.134)	2.65 (.288)	2.80 (.175)	3.37 (.286)
	4	60.0% (13.0%)	17.6% (6.6%)	14.7% (6.4%)	17.6% (6.1%)	49.8% (19.2%)
	3	33.4% (13.2%)	58.1% (13.6%)	46.1% (15.6%)	53.1% (11.8%)	40.4% (14.1%)
	2	5.2% (8.0%)	17.4% (7.1%)	28.9% (13.6%)	21.4% (8.0%)	6.8% (5.9%)
	1	1.4% (1.2%)	6.9% (5.2%)	10.4% (8.6%)	7.9% (6.0%)	2.9% (3.2%)
	Observations	14	14	14	14	14
Non-Muslim countries	Average	3.32 (.222)	2.71 (.241)	2.47 (.234)	2.72 (.245)	3.23 (.226)
	4	45.0% (15.9%)	13.8% (8.8%)	9.7% (5.9%)	15.1% (9.1%)	38.1% (15.6%)
	3	44.3% (12.1%)	51.5% (12.5%)	38.6% (11.4%)	49.2% (11.9%)	49.0% (12.5%)
	2	8.0% (5.0%)	26.9% (9.3%)	40.1% (9.2%)	27.9% (10.4%)	10.4% (6.8%)
	1	2.7% (1.9%)	7.7% (6.7%)	11.5% (8.5%)	7.8% (5.9%)	2.5% (2.4%)
	Observations	105	102	102	103	102

Notes: See Table 3.A.1. Excluded Shia countries are Iran and Iraq.

Table 3.A.6: Robustness: Effect of Democracy without GDP.

Variables	Model C1	Model C2	Model C3	Model C4	Model C5
Intercept	3.10***	2.62***	2.41***	2.59***	3.29***
	(.034)	(.037)	(.038)	(.038)	(.061)
Polity	-.000	-.019***	-.018***	-.009***	-.059***
	(.001)	(.002)	(.002)	(.002)	(.002)
Individual controls	Yes	Yes	Yes	Yes	Yes
-2Loglikelihood	247,442.9	242,404.6	254,163.1	249,968.8	233,044.9

Notes: See Table 3.4.

Summary and Conclusion of Part I

The first part of this thesis has compared and evaluated three approaches to assessing preference heterogeneity: the revealed preference or RP approach, the stated preference or SP approach, and the experienced preference or EP approach. We have also explained their underlying measurement methodologies and identified key differences between the three approaches and types of measures, especially where it regards their psychometric quality (reliability and validity) and usability (as applied in the study of the economics of preference heterogeneity). We have raised many issues, but for this summary and conclusion we limit ourselves to results concerning the comparative usefulness of the three approaches to measuring preferences and assessing preference heterogeneity. The table below summarizes our results thus far.

Table I: Summary of the Comparative Evaluation of the RP, SP and EP Approaches and Measures.

Approach	Revealed Preferences	Stated Preferences	Experienced Preferences
Psychometric quality	● - ●●●●●	●	● - ●●●●
Usability			
Availability	●	●●●●●	●●●●
Scope	●	●●●●●	●●●●
Resolution	●●●●●	●●●●●	●
Explanatory capacity	●	●	●●●●●

Notes: See Table 2.1 and Table 3.5.

Our general conclusion on the three approaches is that there is not one best approach to measuring preferences and assessing preference heterogeneity, either theoretically or practically. The RP, SP and EP approaches all have strengths and weaknesses, which in turn are linked to the particular research question the measures of preference heterogeneity are called upon to address. For the purpose of studying the economics of preference heterogeneity, a first evaluation in Chapter 2 led us to dismiss the RP approach. The reason is that this approach's usability suffers from a lack of explanatory capacity, which is not compensated for by high scores on either availability and/or scope. Chapter 3 elaborated on the idea that SP measures of preference heterogeneity suffer a marginal preferences problem, mistaking marginal preferences, the preference for increasing satiation of an objective given

current levels of satiation, for deep-rooted attitudinal traits, the preference for satiating the objective in general.

Applying the logic of the marginal preferences argument the chapter gives a consistent account of the alleged democracy paradox of Islam: the fact that Muslim-majority countries tend to be less democratic, while both individual Muslims and individuals in Muslim-majority countries tend to have a much more positive attitude towards democracy than other people do. Our empirical analysis shows that what looks like a Muslim democracy paradox is a general phenomenon: the less democracy individuals enjoy, the more favorable their attitude towards democratic political governance. Results therefore support the hypothesis that SP measures, and values surveys in particular, suffer a marginal preferences problem.

Together with the evidence from many other sources, most notably the GLOBE study of 62 societies (House et al. 2004), this casts serious doubt on the psychometric quality (validity) of SP measures of preference heterogeneity. Value preferences or attitudes are a different construct than are marginal preferences. What is more, the marginal preferences problem means that we may wrongfully attribute differences in preferences between groups to, for instance, differences in religious upbringing or ethnic differences, whereas they actually have a rationale in the different circumstances in which these groups find themselves. This, in turn, limits the explanatory capacity of SP measures. Simply put, if values survey scores are driven by contextual factors, they cannot simultaneously explain differences in these circumstances. The evaluation in Table I depicts the low score of the SP approach on these two criteria.

The results of the assessments in Chapters 2 and 3 leave us, partly by elimination, with the EP approach as holding the most promise for useful application in the study of the economics of preference heterogeneity. It should be clear, however, that, as is, a non-negligible part of the EP approach's potential remains unproven. The reason is that not much is known about the psychometric quality of EP measures of preference heterogeneity. This gap needs to be filled and the method's potential assessed further before we can pass a judgment that is more final. We take up this issue in Part II of the thesis to which we turn next.

Part II: Subjective Well-Being and (Heterogeneous) Experienced Preferences

This seeking for a greater measure of welfare and happiness does not indicate a change in values. It is rather a return to values lost in the course of our economic development and expansion.

- Franklin D. Roosevelt, Message to Congress, June 8 1934

Chapter 4

Measurement and Public Policy Uses of Subjective Well-Being[•]

[•] Most of this chapter was written whilst visiting the Statistics Directorate of the OECD. Helpful comments on these parts by Bruno Frey, Jon Hall, Anat Itay, Ramzi Mabsout, and Adolfo Morrone are gratefully acknowledged. An excerpt of some material presented here is published as van Hoorn, A. 2008. A Short Introduction to Subjective Well-being: Measurement, Correlates and Policy Uses. In: OECD (ed.) *Statistics, Knowledge and Policy 2007: Measuring and Fostering the Progress of Societies*. Paris: OECD Publishing, pp. 215-229 (van Hoorn 2008).

4.1. INTRODUCTION

Measures of subjective well-being (commonly abbreviated as SWB) provide crucial empirical input for the experienced preference (EP) approach to measuring preferences and assessing preference heterogeneity. In this chapter we elaborate the SWB construct, discussing how it is defined, measured, and how it may be used, specifically in public policy. We supplement the discussion in this chapter with an extensive empirical assessment of the reliability and validity of SWB indicators and EP measures of preference heterogeneity in Chapter 5.

The measurement of subjective well-being or, in more popular terms, *happiness*, is a contemporary contribution to an age-old debate. Interest in happiness and its pursuit or, more generally, what makes for a good life, started at least two millennia ago with Aristotle (384-322 BC) and, later, Epicurus (341-270 BC). Others, including Thomas Aquinas, David Hume, and Jeremy Bentham continued and added to the debate. The consensus was that happiness matters. In fact, many went so far as to claim that, ultimately, happiness is all that matters.[35] However, massive conceptual difficulties suggest a less strong conclusion: maximizing the experience of happiness, and related notions like flourishing, joy or pleasure, whilst minimizing painful episodes, is the quintessential motive underlying human behavior. Blackburn (1996) and Walsh (2005) give some common philosophical conceptualizations of happiness and well-being.

Given the widespread recognition of the importance of happiness, the study of SWB has a surprisingly short history (cf. Kahneman et al. 1999; Sirgy et al. 2006). Systematic research of SWB, in particular into its determinants, did not start until the second part of the 20th century (Fellows 1966 and Wilson 1967 present early overviews). The literature has bourgeoned since then though. First, psychologists and other social scientists have taken huge steps in their understanding of the factors influencing people's SWB. Argyle (1999), Di Tella and MacCulloch (2006), Diener (1984, 1994), Diener and Suh (1999), Diener et al. (1999), Easterlin (2003), Frey and Stutzer (2002), Headey and Wearing (1992), Layard (2005, 2010), Lyubomirsky et al. (2005) and Nettle (2005) provide overviews of this literature. Second, measurement of SWB has become increasingly sophisticated and it likely continues to do so with the spreading use of and advances in information- and communication technology.

[35] McMahon (2004) gives a short but more elaborate account of the history of the happiness debate. Angner (2005) provides an extensive overview of philosophical issues related to the concept of happiness or, more generally, well-being. King and Napa (1998) present evidence that people indeed attach great value to happiness (see also Diener et al. 1998 and Diener and Oishi 2004).

Paralleling the above-sketched developments, SWB is also increasingly subject of study in economics. The literature is still small, however, especially if compared with that in psychology. As mentioned, this chapter gives a thorough introduction to measures of SWB. We start at the bottom, reviewing the definition of SWB and its hallmark features. We find SWB is mostly an empirical construct and elaborate on the diverse set of measurement instruments and measurement methods that have been developed over the years. The approaches to measuring SWB differ in sophistication and ease of implementation. Depending on the issue one seeks to address and the resources available, an optimal combination may be found. We illustrate (potential) applications of SWB data with a review of their possible uses in public policy. This includes the indirect use of SWB or experienced utility to measure preferences and analyze the costs and benefits of various public goods and bads, but also more direct applications such as the use of SWB to complement existing well-being accounts.[36]

The organization of the chapter is straightforward. Section 4.2 describes the meaning of the term SWB. Section 4.3 discusses some actual SWB scales and methods for obtaining SWB data. Section 4.4 then considers some public policy uses of SWB, including the measurement of preferences. To conclude, Section 4.5 gives final and summarizing remarks.

4.2. WHAT IS SUBJECTIVE WELL-BEING?

In psychology, gradually a consensus has emerged around a working definition of SWB. We can define SWB as "people's positive evaluations of their lives", which, in turn, "includes positive emotion, engagement, satisfaction, and meaning" (Diener and Seligman 2004: 1). As

[36] In case of a direct application of SWB, the interest is in SWB as such: average SWB scores have risen last year; the SWB gap between men and women has decreased; there is a strong cross-sectional relation between GDP per capita and levels of SWB; et cetera. Indirect uses of SWB, in contrast, take the indicators as means to some other end, specifically analyzing issues that do not necessarily have a strong bearing on how well people are doing. Valuation exercises are an example of an indirect use of SWB: with such exercises researchers do not seek to explain variation in levels of SWB on the basis of differences in, say, the concentration of air pollutants, but are interested in SWB and variation therein as a way to put a price tag on this public disamenity (see Chapter 2 and Section 4.4). The measurement of heterogeneity in experienced preferences is a most indirect application of SWB indicators. As we discuss in the next chapter, the direct and indirect use of SWB are associated with separate reliability and validity concerns.

this section shows, however, the concept is somewhat more complicated than that. Two hallmarks of SWB are that (Diener 1994: 106; see also Diener 1984):

1. a positive rating requires not only positive experiences but also the absence of negative ones; and
2. reported scores include a global assessment of how well people think or feel their life is going and is not limited to one or more domains or specific periods.

Looking at the latter two features of SWB in more detail, we see that they suggest that SWB is actually an umbrella term, comprising different elements of people's lives. The SWB literature has broken down people's judgments of their lives along the lines indicated above and into two distinctive components (cf. Andrews and Withey 1976: 18; Diener 2006: 399-400; Diener et al. 1999: 277; Veenhoven 1984: 25). The first component is the *affective* part of overall SWB and refers to both the presence of positive affect and the absence of negative affect. This part is an evaluation guided by emotions and feelings. A *cognitive* part constitutes the second component. This part is a conscious, information-based appraisal of one's life for which people judge the extent to which their life so far measures up to their expectations and resembles their envisioned "ideal" life.

Researchers in the field often make a distinction between self-reported happiness and life satisfaction. Happiness then denotes an emotional evaluation or the affective part of SWB (which in turn must be decomposed in positive affect and negative affect), and life satisfaction the cognitive part.[37] This labeling is not universal, however, and people commonly use happiness and life satisfaction synonymously with SWB. While often clear from the particular context, the ambiguous use of terminology—happiness, life satisfaction, and SWB—shows that the field does not yet agree on any one definition for its over-arching subject matter. Subjective well-being (SWB), however, is a rather more accurate description of the broader field under discussion here.

Economists have recently added to the discussion on SWB, trying to tie it in with more economics-oriented concepts. The work by Daniel Kahneman has been most illuminating. As discussed in Chapter 2, Kahneman draws a distinction between experienced utility, the "hedonic quality" of an outcome, and decision utility, the "weight of an outcome in a decision" (e.g. Kahneman et al. 1997: 375; Kahneman 1999, 2000). The SWB construct and the experienced utility concept match so that in quantitative work measures of SWB are taken as empirical proxies for experienced utility (e.g. Alesina et al. 2004; Di Tella and MacCulloch

[37] See Diener et al. (2010) and Lyubomirsky et al. (2005: 820-822) for extensive discussions.

2005; Frey and Stutzer 2002; Helliwell and Barrington-Leigh 2010a; Luechinger 2009; Rabin 1998; van Hoorn et al. 2010). Following this literature, throughout this chapter and thesis we use the terms SWB and experienced utility interchangeably.

Underlying the definition of SWB as a personal evaluation of one's life is the idea that it is not uncommon for people to make such affective and cognitive judgments (Diener 1994: 106-107). Stated differently, people continuously evaluate their life and some aspects of it, judging which elements are valuable and desirable, and which are not. The affective and the cognitive component of SWB subsequently have a common core but differ in important respects as well. Diener (1994: 107) puts it as follows. On the one hand, "life satisfaction and hedonic level are likely to correlate because both are influenced by appraisals of one's life events, and activities, and circumstances." On the other hand, "life satisfaction and hedonic level are likely to diverge because life satisfaction is a global summary of one's life as a whole, whereas hedonic level consists of ongoing reactions to events (and may also be influenced by unconscious goals and biological factors which may influence mood)" (ibidem; see, also, Lucas et al. 1996 and Diener et al. 2010).

However, measures of SWB do not necessarily refer to life as a whole. It is also possible to investigate separately whether someone is happy or satisfied with different domains of life, such as work, family life, financial situation, or health. Different studies focus on the relation between satisfaction with all relevant aspects of life and global life satisfaction. Andrews and Withey (1976: 127), for instance, find that satisfaction with family, money, amount of fun, housing, national government, and so-called self-efficacy makes up half of all the variance in reported life satisfaction, but the exact relation between domain satisfaction(s) and global life satisfaction is unknown. There can also be substantial heterogeneity in the relation between domain satisfactions and happiness or satisfaction with life as a whole. The values-as-moderators literature on SWB (see Chapter 2) finds, for example, that in wealthier nations, home life satisfaction is a better determinant of satisfaction with life as a whole (Oishi, Lucas et al. 1999) and that for Korean adolescents satisfaction with self matters much less for global life satisfaction than it does for adolescents from the U.S. (Park and Huebner 2005).

To return to the definition of SWB, it is manifest that the field lacks a single, comprehensive definition. Diener et al. (1999: 277) find that "subjective well-being is a broad category of phenomena that includes people's emotional responses, domain satisfactions, and global judgments of life satisfaction" (cf. Diener and Seligman 2004: 1). This is not a universal definition but rather a description of what SWB comprises. The absence of a clear-

cut definition reflects that, as a concept, SWB has no precise theoretical basis (studies like those of Kahneman 1999, 2000 notwithstanding). Indeed, the field appears driven by empirical work, and researchers largely conceptualize SWB through the specific indicators they use, i.e. through the specific questions they ask. The next section discusses the measurement of SWB.

4.3. MEASURING SUBJECTIVE WELL-BEING

The measurement of SWB can be decomposed in three separate elements: (i) the scale or the measurement instrument used, (ii) the method through which researchers elicit SWB reports, and (iii) the final stage of processing data thus collected and reporting the results (or applying them in further research, e.g. to measure experienced preferences and heterogeneity therein). Here we discuss these three components of measuring SWB and how researchers can combine scales and methods in reports of levels of SWB for individuals, specific groups and societies as a whole. The section ends with a brief discussion.

4.3.1. Measurement Instruments

Single-Item Scales

Since the first systematic analyses of SWB began some six decades ago, researchers have developed a large variety of SWB scales and indicators. The most widespread of these are also the most straightforward. Questionnaires simply ask people to state how happy or satisfied they are (with their life as a whole). Two well-known examples of such single-item questions, used in the famous World Values Survey (European Values Study Group and World Values Survey Association 2006; World Values Association 2009), are:

> *Taking all things together, would you say you are…:*
> *1 – "Very happy"*
> *2 – "Quite happy"*
> *3 – "Not very happy"*
> *4 – "Not at all happy"*

and

All things considered, how satisfied are you with your life as a whole these days?

1 2 3 4 5 6 7 8 9 10

Dissatisfied Satisfied

Other single-item questions exist, but these questions reflect the basic approach to measuring SWB and can straightforwardly be used in the estimation of (heterogeneous) experienced utility functions to measure preferences. To address specific issues, such as satisfaction with a domain, the standard approach can be adapted to ask, for instance, "Overall, how satisfied or dissatisfied are you with your job / home life / financial situation?" Most large-scale (multi-country) surveys on SWB include one or more single-item scales like these. The previous chapter already introduced the World Values Survey (WVS), which is the largest cross-country research project from which data are publicly available. In the appendix we give an overview of some of the best-known datasets and their included SWB items, focusing on those that are most readily available. Three of these datasets also provide input for empirical assessments of heterogeneity in experienced preferences in Chapters 5, 6 and 7 of this thesis.

Multi-Item Scales

Since SWB "is a broad category of phenomena" (cf. Diener et al. 1999: 277), one should not expect single-item questions to provide an all-encompassing indicator of people's evaluation of their lives. Multi-item scales may provide a solution and a broad array of such indicators exists (see, for example, Veenhoven 1996: 3-4). Here we discuss two often-used scales, the Positive and Negative Affect Schedule (PANAS) scale (Watson et al. 1988) and the Satisfaction With Life Scale (SWLS; Diener et al. 1985; see also Pavot and Diener 1993a), though many more exist, each with their own strengths and weaknesses.

Watson et al. (1988) developed the PANAS specifically to measure positive and negative emotions. Tailoring the PANAS to include less or more feelings or to be broader or narrower in its cover of time periods is straightforward. It asks about the experience of certain emotional states:

This scale consists of a number of words that describe different feelings and emotions. Indicate to what extent [you felt this way during the past week]. Use the following scale to record your answers:

1 – Very slightly or not at all

2 – A little
3 – Moderately
4 – Quite a bit
5 – Extremely

___ *Interested*
___ *Distressed*
___ *Excited*
___ *Upset*
___ *Strong*
___ *Guilty* [Et cetera]

Investigators can obtain an overall rating by aggregating the answers to the different states, using whichever weighting procedure suits their research question.

Diener et al.'s (1985) Satisfaction With Life Scale (SWLS) is an example of a multi-item measure of SWB that limits attention to a specific aspect of SWB, in this case life satisfaction. The questionnaire for this scale reads as follows:

Below are five statements that you may agree or disagree with. Using the 1 to 7 scale below indicate your agreement with each item by placing the appropriate number on the line preceding that item. Please be open and honest in your responding.

___ *In most ways my life is close to my ideal*
___ *The conditions of my life are excellent*
___ *I am satisfied with my life*
___ *So far I have gotten the important things I want in life*
___ *If I could live my life over, I would change almost nothing*
1 – Strongly disagree
2 – Disagree
3 – Slightly disagree
4 – Neither agree nor disagree
5 – Slightly agree
6 – Agree
7 – Strongly agree

where the answers to these five questions are aggregated with equal weight to obtain the following overall rating:

- *5 - 9* *Extremely dissatisfied*
- *10 - 14* *Dissatisfied*
- *15 - 19* *Slightly dissatisfied*
- *20* *Neutral*
- *21 - 25* *Slightly satisfied*
- *26 - 30* *Satisfied*
- *31 - 35* *Extremely satisfied*

The construct measured by the SWLS resembles the one underlying the single-item life satisfaction question discussed above, though it is obviously more costly to collect both in terms of administration costs and respondent burden.

As with single-item measurement instruments, aggregate scores for both the PANAS schedule and the SWLS provide straightforward input for the EP approach. We note, however, that, as is, not much SWB data obtained from either of these two scales are available, and collection is costly (see also the discussion below).

4.3.2. Measurement Methods

Standard application of single- or multi-item scales of SWB is in one-time, often large-scale surveys (the appendix gives some examples). More advanced methods exist, however. The most notable ones are the Experience Sampling Method or Ecological Momentary Assessment (Scollon et al. 2003 and Stone et al. 1999 provide overviews), and the Day Reconstruction Method, a technique recently developed by Daniel Kahneman and colleagues (Kahneman et al. 2004a, 2004b).

The Experience Sampling Method or Ecological Momentary Assessment

The Experience Sampling Method or ESM is a method to obtain self-appraisals. It asks individuals repeatedly, often over random intervals and for a prolonged period, to make a certain assessment, which can range from filling in questionnaires like the ones just discussed, but can also involve physiological tests (e.g. blood pressure levels, heart rates). The essential

characteristics of ESM are that the evaluations are made frequently and concern the period that has just elapsed. A further feature of the method is that respondents make their evaluation while in their day-to-day surroundings. In fact, these three qualities drove ESM's development in the first place. As Stone et al. (1999: 27-28) point out: (i) studying individuals in their typical surroundings reduces ecological distortions (i.e. has high ecological validity);[38] (ii) on the spot evaluation avoids retrospective distortion; and (iii) high-frequency assessment allows more detailed study of developments over time and the role of changes in circumstances. ESM and Ecological Momentary Assessment or EMA are often equated (Scollon et al. 2003: 7) but Stone et al. (1999) distinguish ESM from EMA on the grounds that EMA also collects information about environmental circumstances while the former limits attention to respondents' self-evaluation.

There are several ways in which to undertake ESM. For the self-appraisal any scale or measure of SWB will do; simply choose the one that best captures the phenomenon or phenomena of interest. Researchers can also organize the actual sampling in many different manners (Scollon et al. 2003: 6). In the early days of ESM, subjects were often responsible for keeping track of intervals between questionnaires, for example. Following technological advances, later on electronic watches pre-programmed to beep at random intervals provided the necessary cues. Nowadays palmtops are increasingly used and these come with important advantages, not least of which are the instant availability of automated questionnaires allowing researchers to analyze completed reports right away.

Unsurprisingly, the benefits offered by ESM and EMA in terms of coverage and ecological validity come at a cost. Using these methods and obtaining regular, virtually online self-assessments is much more expensive than the use of, say, one-time questionnaires, and the final choice of survey instrument requires careful consideration. According to Stone et al. (1999: 30) ESM or EMA is particularly appropriate to measure pain and symptoms of patients.

The Day Reconstruction Method

The Day Reconstruction Method or DRM is another specific approach to collect SWB data. It shares many of the advantages of the other methods but appears easier to implement.

[38] Famous is the phenomenon known as white-coat hypertension: respondents' blood pressure levels are elevated when measured by medical personnel but not when measured at home by respondents themselves (Stone et al. 1999: 27-28).

The technique requires respondents to report about the preceding day and evaluate its various episodes (e.g. hours). Kahneman et al. (2004a) first applied this technique and their questionnaire roughly reads:

> *How did you feel during this episode?*
>
> *Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience.*
>
> - *Happy* *0, Not at all – 6, Very much*
> - *Frustrated/annoyed* *0, Not at all – 6, Very much*
> - *Depressed/blue* *0, Not at all – 6, Very much*
>
> [Et cetera]

Again, however, the method leaves open the use of different measures of SWB.

4.3.3. Processing Subjective Well-Being Data: Aggregate-Level Indicators and Reports

Methods and scales can be combined and processed in different ways. A specific SWB item used in a (one-time) survey typically results in a single number, either the numerically coded answer to the specific question asked or an aggregated index of answers as is the case with either the SWLS or the PANAS scale. This format—individual-level SWB scores for a large sample of individuals—is the required format for applying the EP method to measuring preferences and assessing preference heterogeneity.

SWB data can easily be processed further, however, as is done for studies comparing levels of SWB across countries. Such work takes individual-level SWB data, nearly always measured using a single SWB item (e.g. overall, how happy or unhappy are you with your life on a scale from 1 to 10?), to construct country averages and investigate which countries are the happiest.

Beyond simple surveys and the aggregation of individual answers, ESM/EMA or DRM offer more flexibility because they render multiple observations for a single individual. Using these methods, researchers can bring the raw data together in different ways, choosing the one that best fits the purpose for which the data on individuals' SWB has been collected in

the first place. Notably, researchers can average observations for each individual (and, perhaps, from different SWB items) to obtain a measure of individuals' SWB that is less sensitive to whimsical circumstances (see the second section of Chapter 5). At the aggregate level, the advanced measurement methods offer even more interesting opportunities for constructing measures of SWB for nations or selected classes.

To illustrate, take Kahneman et al.'s (2004b) proposal for "national well-being accounts" (NWBAs). They define national well-being (WB) as a temporal integral of net affect (pleasures minus pains; positive affect minus negative affect):

$$WB = \sum_i U_i/N = \sum_i \sum_j h_{ij}\mu_{ij}/N\,, \qquad (4.1)$$

where U_i is the utility of individual i, h_{ij} is the amount of time he or she is engaged in situation j (e.g. meeting friends, working, et cetera) and μ_{ij} is the net affective experience during situation j (Kahneman et al. 2004b: 432; see also Kahneman 1999, 2000, and Kahneman and Krueger 2006).

Kahneman et al. (2004b) subsequently propose using their own method, namely DRM, to measure net affect over time. Specifically, they favor an approach to the construction of NWBAs that separates the measurement of net affect from observing how individuals allocate their time. The idea is to measure average net affect for a range of situations and activities, then combine this with data from time allocation surveys, followed by a calculation of the temporal integral.

Of course, this is a much more elaborate way of reporting levels of SWB for a particular country than simply aggregating individual answers to the question how satisfied are you. It does show, however, how researchers can combine SWB scales and measurement methods, to give the most complete picture of levels of SWB given the available resources. The idea of a national index of SWB returns in Section 4.4, which discusses public policy uses of SWB.

4.3.4. Discussion

The review above shows there are many approaches to the measurement of SWB, each with their own benefits and drawbacks. The different approaches also share an important commonality: whatever the question or item used, none entails "happiness" or "satisfaction"

in a well-defined form. Answers to SWB items simply reflect how people evaluate their lives or aspects thereof in response to questions asking them, for instance, how happy or satisfied they are. The existence of various SWB measurement instruments (and methods) indicates that the field has not solved the elusive question of what is happiness (nor does it generally claim to have done so). It merely contributes to the age-old research program on the good life and human betterment by analyzing empirical constructs SWB researchers feel have important bearing on this issue. Accordingly, and as alluded to above, the field is characterized by "measurement *before* or *without* theory" to paraphrase a long-standing discussion in economics (see, for instance, Prescott 1986 and, especially, Koopmans 1947). An important question subsequently is whether reported happiness or life satisfaction scores are meaningful. We take up this question in Chapter 5. Here we move on to the discussion of applications of SWB in shaping and appraising public policy and government action.

4.4. PUBLIC POLICY USES OF SUBJECTIVE WELL-BEING

Developments in SWB research—rapidly increasing understanding of the determinants of SWB and continuous improvement of the measuring instruments used—have sparked a small literature discussing and exploring possible uses of SWB in policy making (e.g. Bok 2010; Diener and Suh 1997; Diener et al. 2008; Diener et al. 2009; Dolan and White 2007; Dutt and Radcliff 2009; Layard 2005, 2010; Veenhoven 2002, 2010a). Most famously, the country of Bhutan has declared "gross national happiness" as its overarching policy objective.[39] In this use, happiness appears to be more about the philosophical concept and not so much about the empirical SWB construct, however.

Government policy can usefully apply measures of SWB and SWB research also without it being an explicit objective (cf. Dolan and White 2007). In this use, SWB may guide policy makers because it indirectly informs them about other important features of society. This difference relates to the broad distinction between direct and indirect uses of SWB alluded to in the introduction of this chapter (see Note 36 and the next chapter for an

[39] More information is available from the official website of the Bhutanese government, http://www.bhutan.gov.bt/government/gnh.php or from the Centre of Bhutan Studies' Gross National Happiness website: http://www.grossnationalhappiness.com/default.aspx.

elucidation of this distinction). There is no clear dividing line between the direct and indirect use of SWB in policy making.

From the literature, four concrete policy applications of SWB can be distilled:

1. SWB as a policy goal;
2. SWB as a complement to existing measures of well-being;
3. SWB as a measuring rod in cost-benefit analyses; and
4. the use of SWB in the construction of poverty statistics and equivalence scales.

For the review in this section we have chosen to limit ourselves to applications of SWB in public policy, though it is often rather abstract. We note that many psychologists and psychiatrists apply SWB research every day and the same holds for many people working in medicine or health care (see Fitzpatrick et al. 1992 for a brief review). Furthermore, shared by our four policy applications is the implicit assumption that the goal of governments is to improve well-being. Nevertheless, our survey of public policy uses of SWB is not meant to take a stance on the desirability of welfarism in general, and utilitarianism in particular. Finally, as with all public policy, the use of SWB in shaping and appraising government action faces incentive problems and runs the risk of reducing individuals to statistical numbers (cf. Frey and Stutzer 2010).

4.4.1. Subjective Well-Being as a Policy Goal

Improving Subjective Well-Being: Why and How?

The policy goal of improving SWB constitutes a most direct use of SWB. To have well-being as a policy goal is hardly controversial. Shoving aside the debate of what constitutes well-being, well-being itself is good—if it were not intrinsically good, it probably would not be called well-being. To have promoting SWB as an aim for government policy is much less tautological. The basic argument in favor of SWB as a policy goal is that it captures something that, again, is intrinsically good. Who would not like to experience frequent positive emotions, a minimal amount of negative emotions, and/or lead a meaningful life?[40] Ceteris paribus, most people very likely indeed prefer a life filled with positive affect over a

[40] In the context of policy making, the idea that all well-being is *subjective* (e.g. Sumner 1995) becomes particularly salient. To be sure, the present discussion focuses on SWB defined as "a broad category of phenomena that includes people's emotional responses, domain satisfactions, and global judgments of life satisfaction" (Diener et al. 1999: 277).

life filled with negative affect (cf. King and Napa 1998; Diener et al. 1998; Diener and Oishi 2004).[41]

In the real world, the situation is more complicated and it is not clear how best to go about increasing levels of SWB. At a general level, there are many factors associated with SWB on which governments can exert at least some influence. It goes too far to discuss all factors that might contribute to SWB or at least facilitate the achievement of higher levels of SWB, but democracy, freedom, and quality of governance all correlate positively with average levels of SWB (e.g. Álvarez-Díaz et al. 2010; Bjørnskov, Dreher and Fischer 2008, 2010; Helliwell 2006; Helliwell and Huang 2008; Inglehart et al. 2008; Ott 2010a, 2010b; Veenhoven 2010a; Whiteley et al. 2010). These concepts are rather abstract, however, and somewhat detached from concrete policies governments may implement to foster SWB. The obvious example of a more tangible course of action is to increase expenditure on mental health care (cf. Layard et al. 2007), which is logically expected to have high SWB pay-offs (though, of course, it detracts from other possible causes; see below).

Within the realm of concrete government policies, the most hotly debated policy instrument in the SWB literature is that of taxation. Because of the intensity of the SWB-taxation debate, it serves as a useful case study to see how understanding of SWB and its determinants (i.e. experienced preferences) can help formulate policies, in particular those aimed at fostering SWB. In addition, it serves as an example of a less direct use of SWB; interest is not so much in SWB as such, as it is in people's preferences and in what SWB research can tell us about the efficiency features of taxation.

Taxation, Preference for Relative Income, and Subjective Well-Being: A Case Study

In economics, the welfare consequences of government policies, and taxes in particular, are typically assessed by looking at how they affect people's choices, the so-called behaviorist approach, which has revealed preferences as its empirical building block. Many large macroeconomic models actually contain a small set of core behavioral equations that form the basis for analyses of the costs and benefits of policy alternatives. SWB data and research can add a valuable perspective to this standard behaviorist approach to the study of the welfare effects of a policy (cf. Beshears et al. 2008; Holländer 2001; Gruber and

[41] The ceteris paribus clause deserves emphasis, as the suggestion here is not that SWB is all that matters (cf. Diener et al. 1998).

Mullainathan 2005; Kőszegi and Rabin 2007, 2008; Layard 2010), not least where it concerns the welfare effects of taxation.

The interest in how taxes might be used to promote SWB (or well-being in general) is driven by one of the most striking and robust results in the literature on the determinants of SWB. Empirical analyses consistently show that people not only have a preference for absolute income but also for relative income: both absolute income and how one's income compares with that of others matter for SWB (e.g. Easterlin 1974; Van de Stadt et al. 1985; Clark and Oswald 1996; McBride 2001; Dynan and Ravina 2007; Bjørnskov, Datta Gupta and Pedersen 2008; Inglehart et al. 2008). In SWB terms, individual income can have a negative externality as when income (or consumption) by one individual has a negative effect on other individuals. Luttmer (2005), for example, finds that SWB is harmed by the level of income of one's neighbors, even to the extent that a given-sized increase of a neighbor's income has the same negative effect on one's SWB as a similarly sized decrease in one's own income. Other research presents similar evidence that next to other factors such as health and absolute income, relative income is an important determinant of SWB (see Clark et al. 2008 for an overview).[42]

Results like these on how much people care about relative income and negative (consumption) externalities bring important new insight to the reigning understanding of the welfare consequences of taxation. In particular, authors like Layard (2005), following his earlier work on this issue (Layard 1980), and Frank (1999, 2005) take the evidence on the effect of relative income on SWB to argue that (increased) taxation of labor income can enhance welfare—whereas the common wisdom dictates that such taxation leads to deadweight losses and thereby lowers welfare. Theoretically, the idea is that when consumption of one individual has a negative SWB effect on other individuals, it constitutes a negative externality, much like pollution. Taxing income or consumption then will improve efficiency because it makes the activity deemed polluting, i.e. working for an income, less

[42] Experimental work further confirms that people value relative consumption (e.g. Solnick and Hemenway 1998 2005; Alpizar et al. 2005; Johansson-Stenman et al. 2002). There is also evidence linking relative position to mortality (Wilkinson and Pickett 2006) and social comparisons to patterns of brain activity (Fliessbach et al. 2007). Following Easterlin's (1974) seminal work on the relation between income and happiness, much research has investigated whether in developed countries sustained increases in absolute income lead to higher levels of SWB or whether people in these countries care only about relative income. Recently this debate seems to have intensified (see, for example, Deaton 2008 and Stevenson and Wolfers 2008 and the replies by Easterlin and Angelescu 2009, Easterlin et al. 2010, and Krueger 2008).

attractive in favor of the non-polluting activity, which is to enjoy leisure. Hence, the argument is that rather than introduce distortions and deadweight losses, this tax will improve welfare, at least in terms of SWB.

The argument on how (carefully designed) taxes may be used to increase levels of SWB is the most clearly articulated policy suggestion arising from SWB research. But it is not something that originates strictly from SWB research. Before the use of SWB indicators to measure welfare gained popularity, a large literature already used formal models to scrutinize the policy implications of a utility function with relative income effects; what does it mean for optimal taxation and policies aiming to maximize welfare if people care not just about absolute but about relative income or consumption too? Boskin and Sheshinski (1978), for instance, find that the optimal level of income guarantees and marginal tax rates increases as relative consumption becomes more important in people's utility function. Similarly, Ng (1987) analyzes the implications of relative income effects on the optimal level of public goods provision. Although this level need not be too low when relative income is included in the utility function, the value of public expenditures is likely underestimated as people do not take into account the benefits that flow from public expenditures indirectly through the lowering of private expenditures and the limiting of relative income effects.

Going beyond taxes as an SWB-promoting policy instrument, the above discussion of taxation and its possible role in improving SWB relates to the distinction between direct and indirect uses of SWB. Obviously, SWB research is important in guiding policy makers aiming to improve levels of SWB. At the same time, the value of this research is not limited to suggesting a concrete policy measure (such as a tax) for enhancing SWB scores. Less directly, it can also add an important perspective to existing analyses and established wisdom. By showing that relative income indeed matters, SWB research suggests, for example, that theoretical analyses like those of Boskin and Sheshinski (1978) and Ng (1987) are relevant in real-world policy situations. In this sense, SWB provides the means to uncover areas in which certain policies theoretically argued to be able to affect welfare, may indeed contribute to well-being in ways not previously deemed realistic or feasible.

Below we continue with a discussion of the (direct) application of measures of SWB to complement existing measures of well-being, before moving on to the (indirect) uses of SWB in cost-benefit analyses and the construction of poverty statistics and equivalence scales.

4.4.2. Building a More Comprehensive Account of Well-Being

Many researchers argue that SWB indicators are useful complements to existing measures of how well life is going. Underlying this observation is the growing recognition that, though insightful, there likely is more to well-being than objective measures, alone or in combination, are able to capture. A prominent example of an area in which objective indicators go only so far in giving relevant and meaningful information is the issue of trust. Whether people think they can trust their fellowman or -woman, or whether they expect a "stab in the back" at every possible occasion certainly is an important feature of any society. In addition, it is not clear whether objective indicators can give an accurate and meaningful account of the level of trust in a country. For instance, the number of fraud cases brought before a court per capita per year might tell us something about trust, but few would argue that this number captures the essence of the concept. What is more, a subjective indicator, cleverly asking people how much they trust others, very likely adds insight to existing objective statistics on this aspect of society.

Concerning well-being, the central tenet of SWB research is that its study cannot get around measuring people's cognitive and affective reactions to life as whole or specific domains thereof (Diener and Suh 1997: 200; see also Diener et al. 2008 and Veenhoven 2002). Most significantly, the argument is that SWB is able to capture people's actual experience in a direct manner, while economic, social and environmental indicators do so only indirectly (Diener and Suh 1997: 205). This matters because what is experienced does not have to coincide with objective conditions and indeed often large deviations exist. Crime rates can be low (relatively speaking) but at the same time people may be afraid to wander through the streets alone at night. In fact, researchers often argue that SWB indicators are useful complements to objective indicators precisely because there is a divergence between what people (reportedly) experience and with what is captured in the (existing) objective indicators (see, in particular, Diener and Seligman 2004: 2-3). If SWB correlated perfectly with objective indicators, it would not add anything.

Following the above reasoning, and building on the methods for measuring SWB discussed in Section 4.3, researchers in the field have given the idea of national SWB statistics serious consideration. To give but one example, Diener (2000: 40) proposes to apply experience sampling to a nationally representative sample of respondents (see also Diener and Seligman 2004 and the discussion of Kahneman et al.'s 2004b proposal for national well-being accounts in the third section of this chapter). His national index would record SWB for

a cross-section comprising different age groups, geographical regions, occupational categories and income levels. To maximize the usefulness of SWB thus measured, the actual survey questions would target all components of SWB, the affective component (the presence of positive affect and the absence of negative affect) and the cognitive component. In addition, other constructs thought to capture important aspects of the quality of life and not captured in objective measured could be included, for instance experienced stress and trust.

At a conceptual level, a generally accepted set of guidelines for the construction of national SWB statistics is available (Diener 2006). At a practical level, there is also agreement about the method for collecting the self-reports, ESM/EMA or DRM, quite possibly in some combination, seems the most suitable way to organize the effort (Diener 2006: 402). How subsequently best to carry out the actual collection of the data is still an open issue though. Thousands of people completing ten randomly timed questionnaires on their palmtop each day, complemented with frequent blood pressure and other physiological tests may be the most rigorous approach, but does not seem feasible. In particular, it would be prohibitively expensive and getting a representative sample to subject to this testing for a considerable amount of time (say a year) would probably be next to impossible. Hence, there is a trade-off between rigor and feasibility in the construction of a national index of SWB and how best to deal with this issue is an important challenge for SWB-based complements to existing measures of well-being. Kahneman et al.'s (2004b) proposal for NWBAs that combine net affect measured using DRM with data from time use surveys (again see Section 4.3) seems promising.

In the discussion of this particular policy use of indicators of SWB, it deserves some emphasizing that national SWB statistics need not be an end in and of themselves. The data obtained for the construction of SWB-based well-being accounts may serve other purposes on top of informing government officials and policy makers about how well (different) people in society are doing. Particularly, it would be a valuable search of input for measuring preferences, including the kind of measurement exercises that we discuss next and which are meant to facilitate choosing between various policy alternatives.

4.4.3. Experienced Preferences and Subjective Well-Being as a Measuring Rod

Turning to a less direct use of SWB, an essential aspect in formulating policy is to assess the costs and benefits of alternative courses of action (including doing nothing). Whilst straightforward in principle, practicalities often limit the usefulness and applicability of

formal cost-benefit analyses that compare policy alternatives in terms of a single monetary metric. In many policy areas, public goods and externalities being the ones most familiar to economists, there is no direct way of putting a price tag on the relevant dimensions of the different options available. As briefly discussed in Chapter 2, stated preference (SP) methods can offer an alternative, but it may also be possible to use SWB data and apply the EP approach (cf. Fleurbaey 2009: 1054; Frey et al. 2007, 2010; Stutzer and Frey 2010: 23). Circumventing the problem of missing prices, the evaluation of the effects of public goods, externalities, policy actions, et cetera then takes place in terms of their contribution, negative or positive, to SWB. Valuation exercises that use SWB generally take two steps. The first is to reduce the valuation problem to its contribution to SWB; the second is to compare this contribution with the role income plays in SWB (Adler and Posner 2008; Clark and Oswald 2002).[43] Specifically, the second step consists of calculating by how much the level of income would have to change to offset exactly the SWB effect of the public good, the externality or the policy alternative.

A growing number of studies takes this approach, in which money remains the ultimate "measuring rod" (cf. Pigou 1952), but where the use of SWB indicators helps extend the range of valuation problems the government is able to address through formal cost-benefit analysis. Most popular application is in the valuation of environmental factors such as noise nuisance and air pollution (MacKerron and Mourato 2009; Luechinger 2009; Rehdanz and Maddison 2008; van Praag and Baarsma 2005; Welsch 2006; see Frey et al. 2010 and Welsch and Kühling 2009 for an overview), but the method has been applied in the valuation of other dimensions of life as well.

Frey et al. (2007), for instance, use SWB data to assess the costs of terrorism. Their case concerns the Republic of Ireland during the period 1970-1999. Their analysis attributes differences in reported life satisfaction to variation in levels of terrorism, household income and other personal characteristics. Findings reveal that people, on average, would be willing to give up some 41% of their income to have the level of terrorism reduced to the level prevailing in more peaceful parts of the country.

Another interesting example of SWB-based valuation is the study by Oswald and Powdthavee (2008). They propose a method of using SWB indicators to calculate compensatory claims in tort cases, notably bereavement. The idea is that the positive well-

[43] Kahneman and Sugden (2005) discuss the use of SWB as the direct unit of value in comparing policy alternatives.

being effect of a monetary settlement should offset the emotional damage the perpetrator inflicted on the victim.[44] A typical compensatory amount for bereavement calculated using such hedonic compensation is £100,000.

4.4.4. Poverty Measures, Preferences for Income, and the Construction of Household Equivalence Scales

Though much discussed, poverty is actually rather difficult to measure and many different approaches exist (cf. Citro and Michael 1995; see, also, Foster 1998). Particularly challenging is that poverty appears to have an important relative component. Notably Adam Smith (1776: 1103) already observed that:

> A linen shirt, for example, is, strictly speaking, not a necessary of life. The Greeks and Romans lived, I suppose, very comfortably though they had no linen. But in the present times, through the greater part of Europe, a creditable day-labourer would be ashamed to appear in public without a linen shirt, the want of which would be supposed to denote that disgraceful degree of poverty which, it is presumed, nobody can well fall into without extreme bad conduct. Custom, in the same manner, has rendered leather shoes a necessary of life in England. The poorest creditable person, of either sex, would be ashamed to appear in public without them.[45]

An approach pioneered over 30 years ago by Bernard van Praag and subsequent collaborators uses SWB data to take account of the relative elements of poverty (see van Praag and Frijters 1999 for an overview). For his approach, van Praag used the subjective, survey-based approach to measure the preference for income by way of a utility function that has only income as a factor. Many variations are possible, but the basic question he asked reads as follows (van Praag and Frijters 1999: 418):

> *While keeping prices constant, what after-tax total monthly income would you consider for your family to be:*

[44] The corresponding legal principle is called "restitutio in integrum", which literally means "restoration to original condition" (Oswald and Powdthavee 2008: S218).

[45] See also the literature on preferences for relative income discussed above.

Very bad	$ _____
Bad	$ _____
Insufficient	$ _____
Sufficient	$ _____
Good	$ _____
Very good	$ _____

This is the so-called "Income Evaluation Question" or "IEQ".

Use and analysis of this question has produced two main findings. First, the "utility of income" (van Praag and Frijters 1999: 417) differs over people. Second, subsequent testing revealed a relation between the observed differences and objectively observable characteristics of respondents. In particular, income actually earned is a very significant factor in the preference for (more or less) income: a person earning $40,000 typically thinks $100,000 is a very good income, while a person actually earning $100,000 thinks $250,000 would qualify as a very good income. And the opposite holds as well, someone with a $40,000 income usually thinks a $20,000 income is worse than does a person actually earning $20,000 (van Praag and Frijters 1999: 423).[46]

A government, or in fact, anybody who might be interested in defining poverty in a (partially) relative way can take the IEQ and the idea underlying it as a starting point. A public official may draw the poverty line at the average income deemed very bad by the median household, for instance, but countless definitions are imaginable. The interesting and very useful feature of such a definition is that it does not require (frequent) redefining and that it automatically takes into account the relative aspects of poverty. The definition of a very bad income evolves with the development of average earnings of, say, the median household. Changes therein—increases in times of economic growth and decreases during economic downturns—directly affect the poverty line, and ultimately the poverty statistics.

[46] The sensitivity of answers to the IEQ to actual income earned may appear to be related to marginal preferences problem of SP data (Chapter 3). The fundamental difference is that the IEQ is meant to be sensitive to income earned as a way to establish the importance of income relative to others for individuals' well-being (see the previous note). Alternatively formulated, the IEQ could ask respondents which percentage of their current income they would consider "Very bad", "Bad", et cetera. In this formulation the IEQ would directly capture the "utility of income relative to actual income" and the seeming sensitivity to circumstances, which in turn could indicate a marginal preferences problem, would disappear.

Uses of the IEQ in government policy do not stop at measures of (relative) poverty but straightforwardly extend to the construction of household cost functions and/or equivalence scales (see van Praag and van der Sar 1988). How the needs of households change with their size, i.e. the number of household members, is a long-standing question in economics and an important policy issue. The idea is that the total costs of needs increase with each additional household member, but not in a proportional way; there are economies of scale so that the need for housing space, electricity, et cetera will not be twice as high for a two-person household as it is for a single-person household. In terms of the IEQ, besides income, household size is a second important factor in observed utility functions of income. Keeping other personal circumstances constant, larger households are less satisfied with any given level of income. Knowing the size of the dissatisfaction for households of different sizes subsequently gives an idea of the extent to which households' needs increase with their size, aiding the construction of suitable equivalence scales.

4.5. CONCLUDING AND SUMMARIZING REMARKS

Happiness, pleasure, joy, the good life and related notions have been and continue to be the subject of much debate. They fit the idea that measuring preferences is about finding out what people like and how much they like it. Starting in the second half of the 20th century, the long-standing "happiness" debate—and thereby the possibility of measuring experienced preferences—received an important impetus. At that point, psychologists, traditionally concerned with negative emotions and feelings, slowly became interested in positive aspects of life, for instance in people feeling happy rather than miserable or being satisfied with their life rather than dissatisfied. This research program was highly empirical and developed new indicators to measure how well individuals themselves feel their life is going. The general construct that came out of their efforts is subjective well-being (SWB). SWB refers to a broad set of phenomena that includes positive emotion, engagement, satisfaction and meaning. The key feature of measures of SWB is that they rely entirely on inherently subjective self-appraisals and make no referral to objective circumstances. Many researchers use happiness (or satisfaction) as a synonym for the SWB construct.

SWB is a survey-based measure obtained from self-reports. The basic approach to measuring SWB is fully captured in the idea of asking people how happy or satisfied they are with life as a whole or specific aspects of it. Measurement of SWB has advanced a lot over

the last decades, however. Numerous scales have been developed targeting specific aspects of SWB, for example, positive affect and negative affect or satisfaction with specific life domains such as one's job or one's financial situation. Significant advances have also been made in the method through which data on SWB are obtained. The Experience Sampling Method (ESM) or Ecological Momentary Assessment (EMA), for instance, elicits self-reports repeatedly for a prolonged period, say a week or a month, and generally at random moments during the day. The Day Reconstruction Method (DRM) is a more recently developed technique that shares most of the advantages of ESM/EMA but appears overall to be a more practical method. The technique asks subjects to keep a diary and report on their experience, breaking it down in specific periods, e.g. by hour or by activity (commuting, household chores, work, et cetera). This makes for interesting combinations with survey data on time use. In an important respect ESM/EMA and DRM actually resemble the hedonimeter envisioned by Edgeworth (1881) (see Chapter 2).

Indicators of SWB can have many uses, whichever way they are gathered. In medicine, for instance, it can be used as a screening device to detect psychosocial problems in patients or in clinical trails, where it can help assess the effects of a particular treatment. Another application, of most interest to us, is in the measurement of preferences and group heterogeneity therein. On top of these, SWB indicators have several interesting uses in public policy. With regard to policy making many different applications seem possible. From the literature, several broad areas of application in policy making of SWB can be distilled. Firstly, and most prominently SWB, i.e. increasing its level, can itself be a policy goal. Beyond the goal of improving SWB, indicators of SWB can be used more generally in formulating and evaluating policy. Most notably, SWB data can give public policy makers relevant information about the costs and benefits of their policies and be used to assess the effects of alternative courses of action. Following the use of SWB to measure preferences, a most interesting application of SWB indicators is in the valuation of environmental externalities such as noise nuisance and air pollution. In a two-step procedure the SWB effect of public goods and bads can be compared to the contribution income makes to people's self-reported happiness or satisfaction to then calculate the monetary amount that would exactly offset the adverse or beneficial effects of the public (dis)amenity. Finally, SWB can be used in the construction of poverty statistics that take the relative features of poverty into account and in the construction of household equivalence scales. Although SWB is not widely applied in shaping and appraising policy, it is clear that policy makers may find measuring SWB useful. What is more, as SWB research becomes more sophisticated—we are thinking mostly about

the Day Reconstruction Method—its usefulness for public policy making is set to improve further.

As with all societal indicators, important caveats apply to SWB indicators and their use in formulating government policy as well. Most importantly, SWB research or measures of SWB do not provide a policy recipe. Although we are unaware of any indicator for which changes—up, down, or lack thereof—do come with straight-cut policy advice, this point is well emphasized. Thus, the SWB literature may ultimately have an effect on what governments do or do not do and how they do it, but the field itself only hands the tools and insights, not a roadmap.

More generally, an important open question is whether SWB indicators are meaningful, i.e. reliable and valid. Minimum levels of reliability and validity are a prerequisite for any use of indicators of SWB, whether direct, in assessments of how well individuals or societies are doing, or indirect, as when SWB data is used to measure preferences and assess preference heterogeneity. We take up this issue in the next chapter, which adds to the introduction to SWB given in this chapter and completes our evaluation of the usefulness of the EP approach for the study of the economics of preference heterogeneity.

APPENDIX CHAPTER 4: SURVEYS WITH SUBJECTIVE WELL-BEING ITEMS

World Values Survey (http://www.worldvaluessurvey.org)

The World Values Survey (WVS), which we already used for our analyses in the previous chapter, has included many different SWB items. Examples, next to the ones mentioned in the body of the paper, are:

- We are interested in the way people are feeling these days. During the past few weeks, did you ever feel [Particularly excited or interested in something]: 0, No – 1, Yes.
- Please look at this card and tell me for each word, how often you feel this way at home? [Relaxed]: 1 "Often" – 2 "Sometimes" – 3 "Rarely" – 4 "Never".
- Overall, how satisfied or dissatisfied are you with your job?: 1 "Dissatisfied" – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 "Satisfied".

We use the WVS again in Chapters 5 and 6.

European Social Survey (http://www.europeansocialsurvey.org)

The European Social Survey (ESS) is a biennial survey of some 30 European countries. Complete results from the first three rounds (2002/2003 2004/2005 and 2006/2007; 30,000 face-to-face interviews each) are publicly available and so are partial results for the fourth wave. The ESS includes two general SWB items:

- Taking all things together, how happy would you say you are?: 0 "Extremely unhappy" – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 "Extremely happy".
- All things considered, how satisfied are you with your life as a whole nowadays?: 0 "Extremely dissatisfied" – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 "Extremely satisfied".

In the same format the ESS asks respondents how satisfied they are with certain aspects of their society, such as their national government. The above two items are the same items that we will use in our empirical analysis of the psychometric quality of EP measures of preference heterogeneity in the next chapter.

Socio-Economic Panel (http://www.diw.de/en/soep)

The German Socio-Economic Panel (SOEP or GSOEP) follows a panel of individuals over time, interviewing them yearly. Running since 1984, most individuals included in the survey, nowadays some 20,000 residing in 11,000 households, are interviewed more than once. This makes it possible to study variation in SWB both between individuals and within an

individual over time. The dataset is available for a small fee (€ 30). Measures of SWB included focus on (domain) satisfaction, for example:

- Satisfaction With Life Today. Satisfied: On Scale 0-Low to 10-High: 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10.
- Satisfaction With Health. Satisfied: On Scale 0-Low to 10-High: 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10.

We use data from the SOEP in Chapter 7 where we analysis differences in preferences between (former) East and West Germans.

General Social Survey (http://www.norc.uchicago.edu/GSS+Website)

The General Social Survey (GSS) interviews U.S. citizens concerning, amongst others, their behavior, attitudes, and SWB. The survey started in 1972 and included SWB items are:

- Taken all together, how would you say things are these days -- would you say that you are very happy, pretty happy, or not too happy?
- Now I'm going to read a list of different feelings that people sometimes have. After each one, I would like you to tell me on how many days you have felt this way during the past 7 days. On how many days in the past 7 days have you [Felt sad]?
- We are interested in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all?
- All things considered, how satisfied are you with your family life?: Completely satisfied 1 – Very satisfied 2 – Fairly satisfied 3 – Neither satisfied nor dissatisfied 4 – Fairly dissatisfied 5 – Very dissatisfied 6 – Completely dissatisfied 7.
- How much of the time during the past 4 weeks [Have you felt downhearted and blue]? All of the time 1 – Most of the time 2 – A good bit of the time 3 – Some of the time 4 – A little bit of the time 5 – None of the time 6.

Gallup World Poll

Gallup's World Poll is the most comprehensive survey, probing respondents concerning their SWB, covering over 140 countries. Researchers can buy the data from Gallup for a substantial fee, so that the use of the World Poll is limited to those who can afford it. The World Poll's questionnaires use several SWB items, including the following:

- All things considered, how satisfied are you with your life as a whole these days? Use a 0 to 10 scale, where 0 is dissatisfied and 10 is satisfied.
- Did you experience [Depression] during a lot of the day yesterday?
- Now, please think about yesterday, from the morning until the end of the day. Think about where you were, what you were doing, who you were with, and how you felt. Did you smile or laugh a lot yesterday?
- Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. Suppose that we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?
- Now, would you tell me how you feel about various aspects of your life today? Would you say you are very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with your current housing, dwelling, or place you live?

Chapter 5

Psychometrics for Subjective Well-Being and Its Use in the Measurement of Preference Heterogeneity: An Empirical Assessment

5.1. INTRODUCTION

Our discussion of subjective well-being—what it is, how it is measured, and how it may be applied—in the previous chapter has laid the foundation for our substantive interest in indicators of SWB, namely as input for applying the EP approach and thus providing valuable means for measuring preferences and assessing preference heterogeneity. This chapter deals with the reliability and validity issues that arise in the application of SWB data for these purposes. It presents a detailed psychometric assessment, for which we distinguish between direct and indirect uses of SWB (cf. the previous chapter).

Distinguishing between direct and indirect applications and addressing the psychometric issues that arise in both contexts separately serves as an added quality check. In case of a direct use of SWB, we only need to ascertain the reliability and validity of indicators of SWB to ensure that we can trust the outcome or conclusion of our analysis. As long as SWB indicators are sufficiently reliable and valid, it makes sense, for instance, to compare men and women and to draw (quantitative) conclusions concerning their respective levels of SWB. We may argue, in contrast, that indirect uses of SWB, not least of which is the measurement of preference heterogeneity, require more assurance. That is, they may reach a point at which the reliability and validity of SWB indicators no longer satisfy as an assurance that the study of the indicators actually renders meaningful results. The logic of this derives from the nature of the findings of the analysis, which differs between direct and indirect uses. In direct applications, what comes out of the analysis remains closely linked to the original SWB data—the SWB gap between men and women has decreased, SWB levels have risen last year, SWB is lowest for middle-aged people, et cetera. The more indirect the application, the looser this link is. SWB-based valuation exercises (Carroll et al. 2009; Frey et al. 2009, 2010; MacKerron and Mourato 2009; Luechinger 2009, 2010; Luechinger Raschky 2009; Rehdanz and Maddison 2008; van Praag and Baarsma 2005; Welsch 2002, 2006, 2007, 2009; Welsch and Kühling 2009), for example, draw their conclusions in monetary terms: a given concentration of pollutants lowers SWB by so much that incomes would need to increase by so many dollars or euro's annually to compensate. EP measures of preference heterogeneity are even more detached from actual SWB indicators. We take the distance of such empirical results to the original measure of SWB, the psychometric quality of which needs proving first, important enough to raise additional validity and reliability concerns.

Our solution, as indicated, is to perform a further quality check in the form of additional psychometric evaluation. Applied to the measurement of heterogeneous

preferences this means we do not just assess the reliability and validity of the SWB indicators involved but also assess the reliability and validity of the measures of preference heterogeneity obtained using SWB as empirical input. This chapter therefore deals with the psychometrics of SWB indicators and their use in the measurement of preference heterogeneity in two steps. Recognizing that the reliability and validity of indicators of SWB is a prerequisite for their application in the measurement of (heterogeneous) experienced preferences, we first assess the reliability and validity of SWB. We do so by way of a literature survey. The psychometric quality of SWB indicators has been extensively studied and this body of research provides enough information to draw conclusions on the reliability and validity of indicators of SWB. For the second part of our assessment we cannot rely on existing literature. The reason is simply that almost nothing is known about psychometric quality in the context of indirectly applying SWB data including the measurement of preferences. We therefore conduct, as alluded to above, an extensive psychometric evaluation of EP measures of preference heterogeneity.

The way we actually go about evaluating the psychometric quality of EP measures of preference heterogeneity is to look at the consistency of measured preference heterogeneity across different types of SWB indicators (cf. Finkelstein et al. 2008). In the previous chapter we mentioned that many different indicators of SWB exist and that each of these may be associated with different components of the overarching SWB construct. Specifically, SWB has an emotional and a cognitive component, where self-reported happiness is associated more with the former, and satisfaction with life more with the latter (see, for example, Diener et al. 2010). Using both happiness and satisfaction scores, our analysis finds that measured preference heterogeneity is largely independent of the particular SWB indicator used: the country pattern in preferences measured using self-reported happiness replicates itself in the pattern obtained using life satisfaction to a considerable degree. We find that correlations between different EP measures of preference heterogeneity can reach up to 0.95, which bodes well for the reliability and validity of these measures. Importantly, psychometric quality appears strongly improved by having a sufficient amount of data available.

From the results of existing evaluations brought together in this chapter, we conclude that measures of SWB are reasonably reliable and valid, allowing many different kinds of uses of SWB data. Similarly, based on the high level of consistency between preferences measured using different types of SWB indicators we conclude that the EP method succeeds in rendering meaningful measures of preference heterogeneity.

The remainder of this chapter starts with a survey of existing literature to assess whether indicators of SWB are meaningful, expressly looking at their reliability and validity, in Section 5.2. Section 5.3 then presents an extensive empirical assessment of the reliability and validity of EP measures of preference heterogeneity. Using two large-scale datasets we show that measured heterogeneity tends to be highly consistent across different indicators of SWB. We analyze the robustness of our empirical results in Section 5.4, finding further support for the reliability and validity of the EP approach and SWB-based measures of heterogeneous preferences. Section 5.5 concludes with some final and summarizing thoughts.

5.2. ARE MEASURES OF SUBJECTIVE WELL-BEING RELIABLE AND VALID? A SURVEY

As mentioned in the introduction to this chapter, for direct uses of SWB to render meaningful findings and conclusions it suffices that measures of SWB are reliable and valid. SWB researchers have done a good deal of work investigating the reliability and validity of their measures. Here we give an excerpt of some of the most prominent findings in this literature and add a discussion of important related findings. Further surveys of assessments of the psychometric quality of SWB indicators can be found in Diener (1984, 1994), Diener et al. (1999), Diener et al. (2009: 67-94), Frey and Stutzer (2002), Headey and Wearing (1992: 25-38), Helliwell and Barrington-Leigh (2010a), Nettle (2005), and Layard (2005), amongst others.

5.2.1. Evidence on Reliability

The reliability of SWB indicators has mostly been assessed by looking at test-retest correlations. This research has found that minor differences in circumstances and technical features of the specific questionnaire used can have significant influence on reported levels of SWB (see Schwarz and Strack 1999: 62). Correspondingly, the test-retest correlation for single-item measures often does not exceed 0.60 when the same question is asked twice during a one-hour interview (ibidem; see also Andrews and Withey 1976). Kahneman and Krueger (2006: 7) report a test-retest correlation of 0.59 in a sample of 218 respondents interviewed and re-interviewed two weeks apart. Veenhoven (1996: 6) discusses the effect of the lag between times of asking on test-retest correlations. If respondents are asked twice in

the same interview correlations are approximately 0.70; if the lag spans a week, test-retest correlation drops to about 0.60. He points out, however, that people seldom make large moves, e.g. from "satisfied" to "dissatisfied". Finally, a detailed study by Ehrhardt et al. (2000) examines panel data for Germany for the period 1984-1994. They find that at the start of the sample period the year-to-year correlation was 0.45 but that, for unknown reasons, it gradually increased to 0.54. At the same time, however, the correlation between SWB reported in 1984 and that reported in 1994 was only 0.29. This suggests that the measure is reliable in the sense of having a structural nature. At the same time, and more importantly, it shows that SWB is susceptible to change as well, for instance in response to changes in objective circumstances such as income or employment status (cf. Veenhoven 2010a).[47]

In general, the reliability of SWB measures as evidenced by test-retest correlation is lower than that found for common microeconomic variables such as personal income (Krueger and Schkade 2008). It appears to be more similar to the reliability of other subjective measures such as self-reported health (see, for example, Crossley and Kennedy 2002; Stewart et al. 1988; VanderZee et al. 1996). Next to the lag between times of asking (for example one hour or two weeks), the specific measurement scale used also affects test-retest correlation, however. Studies show that the more advanced measures, such as multi-item scales, produce more reliable SWB scores (see the overview in Krueger and Schkade 2008). In addition, indicators targeting the affective part of SWB appear to have lower test-retest correlations than indicators aiming to measure the cognitive part. Thus, measures of happiness tend be more susceptible to whimsical circumstances—the outcome of a soccer match and the weather, whether it is rainy or sunny, are famous examples (Schwarz and Strack 1999: 62)—than are measures of life satisfaction (Krueger and Schkade 2008).

[47] Compare this with SP measures, which often seek to capture value dispositions or attitudes. These are typically thought of as enduring traits, so that test-retest correlations are expected, perhaps even required, to be much higher than is the case with SWB. The evidence on the sensitivity of SP measures to changes in circumstances (e.g. Clarke et al. 1999; Duch and Taylor 1993), of course, suggests that the reliability may be low. In fact, as the empirical case study in Chapter 3 has shown, the reliability of the measures may be so low, i.e. so sensitive to circumstances, that the measures fail to capture deep-rooted value traits altogether and elicit marginal preferences instead.

5.2.2. Evidence on Validity

The many available reviews (see above) testify to the attention SWB researchers have paid to the validity of their key empirical construct. The conclusion of the reviews is that indicators of SWB do quite well on various validity tests mentioned in the Chapter 5. Table 5.1 presents a brief overview of some commonly cited evidence on the validity of measures of SWB.

Table 5.1: Validity of Measures of Subjective Well-Being.

Study	Main findings
Andrews and Crandall (1976)	Different single-item measures of SWB give similar results and of the total variance in single-item scales, about 64% is valid variance.[48] Validity can be increased further by using composite multi-item indicators; a five-item scale would typically have about 80% valid variance.
Costa and McCrae (1988)	Self-reported SWB scores show convergence with reports filed by spouses.
Diener et al. (1991)	Self-reported levels of SWB converge on those reported by others (e.g. by family and friends).
Ekman et al. (1990)	Subjective reports on the experience of positive emotions during an experiment correlated with the number of genuine smiles (referred to as "Duchenne" smiles) that occurred during the experiment.
Kammann et al. (1984)	Factor analysis shows that 13 different SWB scales measure the same general well-being factor. In addition, the negative halves of these scales overlap with the construct measured by scales of neuroticism, depression and so-called trait anxiety.
Larsen et al. (1985)	The authors consider the validity of various SWB measures and report strong negative relations between positive overall SWB on the one hand and neuroticism and self-reported symptoms on the other; and strong positive correlations between overall SWB and ratings for several life domains, such as friendship, love life and financial situation.
Pavot and Diener (1993b)	Self-reported SWB correlates highly with peer reports, the personality traits of extraversion and neuroticism but not with current mood. Context has a slight effect on reported scores when single-item measures are used but multi-item scales appear largely immune.
Rodgers et al. (1988)	A comparison of SWB scores obtained using different indicators shows that more than 50% of total variance is valid variance and directly related to the construct of interest while only 10% is attributable to method (and one third is due to measurement error in the specific indicator used).

[48] The difference in SWB as actually perceived by the different respondents constitutes valid variance. Other (non-valid) variance results from measurement error, for instance, from respondents entertaining a different interpretation of the response categories (being "very happy" does not mean the same thing to everybody) (cf. Andrews 1984: 413).

Table 5.1, continued.

Study	Main findings
Seidlitz and Diener (1993)	For people with higher SWB scores, memories of happy events are more accessible than for people with lower SWB, while current mood has only a modest effect on the availability of memories.
Watson and Clark (1991)	There is a strong correlation between self-reported scores on eight different negative affect and positive affect scales and ratings by peers.

The apparent pervasiveness of research validating measures of SWB notwithstanding, the indicators have also received much criticism. Very often this criticism is based on the above-mentioned finding that whimsical (and relatively minor) circumstances like the outcome of a soccer match have an impact on the level of SWB people report (Schwarz and Strack 1999: 62). The errors generally seem to be of a random rather than a structural nature, however, working both to increase and decrease SWB scores. Consequently, the measuring instrument is unlikely to be systematically biased. Indeed, using large enough samples would go a long way in addressing possible problems introduced by contextual factors influencing the reported level of SWB (cf. Di Tella and MacCulloch 2006: 29; Frey et al. 2009: 320; Krueger and Schkade 2008: 1843).[49] And this is an advantage of ESM/EMA or DRM: using repeated measures of SWB for a single individual lessens the influence of minor life events. In addition, a recent experiment finds that SWB scores are largely robust to priming with important life events such as bereavement or parental divorce (Sgroi et al. 2010).

Another challenge SWB indicators face concerns their validity in cross-national comparisons. Here the issue is the possible impact of culture and language on SWB ratings (see Diener and Suh 2000, Diener and Oishi 2004, Oishi 2010, Veenhoven 2010b, and Wierzbicka 2004 for introductions to these issues). By the very nature of SWB and the way it is measured, i.e. through questionnaires, cultural and linguistic factors can affect SWB ratings. As Wierzbicka (1999: 31) puts it, "feelings represent the subjective experience of biological (physiological) events, but they can be categorized in a variety of ways, and they are categorized differently in different cultures". SWB may thus be susceptible to influences

[49] Boulding (1972: 466) describes this effect of the law of large numbers as follows: "Even if we cannot get very reliable measures for the individual, by the famous principle which I have sometimes called "Katona's Law", that the summation of ignorance produces knowledge, we may find an operation or instrument with self-cancelling random factors which will give us a much better measure for a hundred or a thousand individuals than we can get for one." Recall that a bias only matters for measured heterogeneity if it differentially affects the groups included in the comparison (see Section 2.4).

that differ systematically between countries, meaning that measured levels of SWB can exhibit a structural bias.

Evidence on this issue is not overwhelming, however. Language generally does not seem to pose much of a problem. Some tentative evidence comes from Inglehart (1990: 28-29). He looks at average levels of self-reported life satisfaction for German-speaking, French-speaking and Italian-speaking Swiss respectively, and finds that for all Swiss average scores are higher than the scores of their linguistic counterparts, citizens from Germany, France and Italy. More convincingly, Ouweneel and Veenhoven (1991) compare self-reports across language groups in bilingual nations and do not find a reliable language effect. In addition, they find that the percentage of "don't know" answers to SWB questions (ignorance) does not differ much between countries.

Cultural characteristics may be a more important factor and possible source of bias. Uchida et al. (2004) review the cross-cultural SWB literature and find cultural dissimilarities. More concretely, Scollon et al. (2004) examine three different types of global self-reports of pleasant/unpleasant emotions for people from five groups each with different cultural backgrounds. They find that for all three indicators culture is a source of response differences between the groups. Vittersø et al. (2005) find a similar result in their analysis of life satisfaction data from Norway and Greenland. Overall, no clear picture emerges, however, because, for example, Diener et al. (1995) discard a cultural explanation for particular observed differences in SWB (even with income controlled for). Notably, humility and social desirability (where people give the answers they think others will regard favorably; see also Section 2.4) do not affect reported SWB scores, leaving SWB differences between Pacific Rim countries and the U.S. unexplained. In addition, it appears that neither ignorance, nor a general lack of interest, nor a bias towards neutral and conforming answers drives the observed differences.

Furthermore, the simple fact that culture is a determinant of SWB doe not imply that the measure is somehow biased. In contrast, the SWB effect of culture may be genuine (e.g. Veenhoven 1987). This latter conclusion also holds, more strongly in fact, for research on cultural variation in the predictors of SWB and heterogeneous experienced preferences (Alesina et al. 2004; Aknin et al. 2010; Clark et al. 2005; Diener and Diener 1995; Diener et al. 2000; Kuppens et al. 2008; Kwan et al. 1997; Malka and Chatman 2003; Oishi, Lucas et al. 1999; Oishi, Suh et al. 1999; Park and Huebner 2005; Schimmack et al. 2002; Suh et al. 1998). Cultural factors are not a source of bias in the empirical analysis of the structure of SWB in terms of its predictors—a statistical error or nuisance we may want to dispose of.

Rather, they reflect important and valid differences in how much certain cultural groups or societies enjoy various aspects of their lives, either emotionally or cognitively or both.

Related research has looked more directly at emotions that are either shared by, or unique to, different cultures. For instance, Ekman et al. (1987) find that there is cross-cultural agreement on the type of emotions expressed in facial images (sadness, joy, et cetera). Results from this line of research indicate that there are universal basic emotions, and despite some methodological critique (Ortony and Turner 1990), it seems that ""human emotions" vary a great deal across languages and cultures," but that "they also share a great deal" (Wierzbicka 1999: 34). This similarity of emotions, in turn, suggests that indicators of SWB do have meaning across national, linguistic and cultural borders.

Finally, some culturally rooted "bias" is unavoidable in many indicators, even, for instance, in the comparison of national income numbers. Despite a common definition, habits and customs regarding what activity takes place through the market do affect these statistics (cf. Frankel 1953; Kuznets 1972; Morgenstern 1975; Sen 1976). Rosen (1996: 734-735) goes to the heart of the problem where he notes that "[i]n Sweden a large fraction of women take care of the children of women who work in the public sector to care for the parents of the women who are looking after their children." Again here culture's influence may be seen as a source of error or as a genuine effect warranting further research.

5.2.3. More on the Meaning of Subjective Well-Being

The findings from the SWB literature presented above support the conclusion that SWB indicators are meaningful. There is less evidence about the meaning of these indicators, other than that they are capturing something important, however. SWB remains largely an abstract concept. Some additional research goes some way in tackling this issue.

By now there is a substantial body of research that relates SWB to human physiology and brain activity. Levesque et al. (2003), for example, applied functional magnetic resonance imaging (fMRI) to analyze the neuroanatomical correlates of sad feelings in healthy children. Measured brain activity showed sad feelings are associated with significant bilateral activations of the midbrain, the medial prefrontal cortex, the anterior temporal pole and the right ventrolateral prefrontal. Ryff et al. (2004) present some preliminary findings showing that people with higher SWB (specifically, more meaning, purposeful engagement, et cetera in their lives) have lower levels of daily salivary cortisol and pro-inflammatory cytokines. In addition, for them the duration of REM sleep is longer than for those with lower levels of

reported SWB. Urry et al. (2004) administered SWB questionnaires to people prior to analyzing their brain activity. A higher level of reported SWB was associated with greater left than right superior frontal activation. Steptoe et al. (2005) report that positive affect is associated with reduced neuroendocrine, inflammatory and cardiovascular activity. Positive affect also has a negative relationship to cortisol output during the day (controlling for other factors such as age and gender) and heart rate. During mental stress testing in the laboratory, people with higher positive affect had smaller plasma fibrinogen stress responses. Finally, a study by Rainville et al. (2006) shows that basic emotions are associated with distinctive patterns of cardiorespiratory activity.

This and other work on the physiological aspects of SWB tells us that SWB is not merely a construct that exists within people's minds. Rather, variations in SWB manifest themselves in observable biophysical phenomena as well as in people's self-reports. One can judge indicators of SWB meaningful, for instance, because they correlate with objective variables that have a solid biological foundation.

5.2.4. Discussion

The evidence presented in this section suggests that many measures of SWB are meaningful and capture important information. As Diener and Suh (1999: 438) conclude, SWB indicators "have a degree of validity and are often not as contaminated as popular lore might suggest." On top of this, the field is bound to advance, resulting in ever-more reliable and valid measures (see also the previous chapter). An interesting possibility is that the SWB construct will be increasingly connected to objective biological phenomena such as patterns of brain activity. Such a development would actually fit the growing trend of merging social sciences with cognitive and brain sciences, as with the developing field of neuroeconomics (Bernheim 2009; Camerer 2007; Camerer et al. 2004, 2005; Glimcher and Rustichini 2004; Rustichini 2005). Although our current understanding of the neuroanatomical correlates of SWB and the mechanisms underlying the observed relations between SWB and brain activity is sketchy, knowledge in this area advances at great speed. Ultimately, we might actually be able to "calibrate" SWB scores with a mixture of physiological phenomena such as hypertension (cf. Blanchflower and Oswald 2008) giving it an objective anchor. To give a very crude illustration of this future possibility, take a scale running from 1, dissatisfied to 10, satisfied (see Section 4.3). People from country X reporting a 9 on this scale may have an average "brain activity level" of 100 while the average "brain activity level" of people from country Y

with the same score (9) is 110. Vice versa, people with a "brain activity level" of 100 in country X on average report a 9 on the dissatisfied-satisfied scale while in country Y people with this "brain activity level" (100) on average report an 8 on the same scale.

5.3. EMPIRICAL ANALYSIS OF THE PSYCHOMETRIC QUALITY OF HETEROGENEOUS EXPERIENCED PREFERENCES

In the introduction to this chapter we briefly expanded on the value of a supplementary psychometric quality check in cases of indirectly applying measures of SWB. In this section we apply this logic to the use of SWB data in the measurement of preferences and heterogeneity therein. As noted in Chapter 2 and the conclusion to Part I of this thesis, up to this point, the EP approach to measuring preference heterogeneity lacks evidence on the reliability and validity of its measures. Here we fill this gap.

Finkelstein et al.'s (2008: 25-26) study is the only one we are aware of that provides a quality assessment of heterogeneous experienced preferences, though it is highly implicit. They compare experienced preferences estimated using different indicators of SWB to find that they render comparable results concerning the health-state dependence of the utility function. We build on their study and examine the consistency of group variation in the experienced utility function across different indicators of SWB. The idea is that meaningful measurement of heterogeneous experienced preferences requires that results obtained using different SWB indicators are similar, whatever the exact nature of these findings. More concretely, we find that the extent to which group variation in experienced preferences measured using one SWB indicator replicates itself in group variation in experienced preferences measured using another SWB indicator tells us whether group variation in experienced preferences is of a structural rather than a random nature. If it turns out that heterogeneity in experienced preferences has a structural component that is independent of any specific SWB indicator, this bodes well for the reliability and validity of experienced preference measures; these measures seem able to capture meaningful differences between groups.

Our actual assessment of the psychometric quality of experienced preference measures starts with estimating SWB functions for a large number of collectives identified beforehand. We use countries as our units of analysis as national boundaries likely delineate cultural heritage associated with differences in preferences (cf. Hofstede 1980, 2001; House et al.

2004; Inglehart 1990, 1997; Inglehart and Baker 2000; Schwartz 1994; Schwartz and Sagiv 1995). We assess heterogeneity by comparing the coefficients of the estimated SWB functions for the various nations. Consistency is examined by repeating this exercise for different SWB indicators and assessing the extent to which heterogeneity found using one indicator replicates itself in heterogeneity found using another indicator. Formally, we calculate the correlation between coefficients obtained with different SWB indicators across the countries included in our analysis. The section includes two separate assessments, each one drawing on a different large-scale multi-country survey of SWB and individual socioeconomic and sociodemographic characteristics. We first describe our method and data, however. Later on, in Section 5.4, we check the robustness of our results.

5.3.1. Method and Data

General Approach

Our method of assessing the consistency of heterogeneity in experienced preferences is straightforward. We take two different indicators of SWB, namely self-reported happiness and satisfaction with life, and empirically analyze whether they render a similar picture of group variation in experienced preferences (cf. Finkelstein et al. 2008). For this purpose, we relate these indicators to a broad range of (possible) determinants of SWB, separately for different collectives. In the identification of the collectives for which we compare experienced preferences, different individual-level criteria can be used. We take countries as the relevant groups as this is the most common aggregate-level unit of analysis and renders a large number of classes for which to compare experienced preferences. Other criteria may be applied just as easily, however: left-wingers versus right-wingers, religious versus non-religious individuals, et cetera (cf. Alesina et al. 2004; Di Tella and MacCulloch 2005; Di Tella et al. 2010; Lelkes 2006).

As discussed in the previous chapter, happiness and satisfaction are associated with the two different components of SWB: happiness links to the affective part and involves an emotional assessment, whilst satisfaction involves a cognitive evaluation (e.g. Diener et al. 2010). Since these two indicators target different aspects of people's self-assessed well-being, there is no reason to assume that preference heterogeneity measured using one indicator will coincide fully with preference heterogeneity measured using the other indicator. If EP measures of preference heterogeneity are meaningful, i.e. reasonably reliable and valid, we do expect some consistency between the measures obtained using self-reported happiness as the

proxy for experienced utility and those obtained using satisfaction, however. At the least, we expect preference heterogeneity measured using either of these indicators to have some consistency—and not be contradictory. If the data do not show such a match between group variation in experienced preferences measured using self-reported happiness and group variation in experienced preferences measured using self-reported satisfaction it would seem that the EP approach does not render particularly reliable or valid measures of preference heterogeneity.[50]

For guidance on which factors to include in our experienced utility functions we turn to the large literature on the determinants of SWB in both psychology and economics. This body of research has identified some "usual suspects," (situational) factors and practices widely established as significant determinants of SWB (see Argyle 1999; Di Tella and MacCulloch 2006; Diener 1984, 1994; Diener and Suh 1999; Diener et al. 1999; Easterlin 2003; Frey and Stutzer 2002; Headey and Wearing 1992; Layard 2005, 2010; Lyubomirsky et al. 2005; Nettle 2005). We include many of these in our two assessments. A restriction is that we only include individual-level determinants of SWB that are available in more or less similar form in the two datasets we use in our analyses. This criterion can be limiting, but our assessments include all factors typically included in empirical analyses of SWB and its determinants. On top of that some additional factors have been added. The independent variables concern, amongst others, income, marital status, health, and employment status. By including a broad set of determinants our analyses render a comprehensive assessment of the consistency of EP measures of preference heterogeneity across a wide range of practices and situational factors.

Empirical Strategy

Both our psychometric evaluations take two empirical steps. We first assess the relation between our two measures of experienced utility and individual determinants separately for each independent variable and each country. Formally, we start with estimating an empirical model of the following general form:

[50] We take the consistency of measured preference heterogeneity across SWB indicators to capture both reliability and validity and hence psychometric quality. Note, however, that consistency is somewhat closer to reliability than to validity (see Section 2.4): in our analyses, we measure constructs that are distinct—happiness-based versus satisfaction-based experienced preferences—but that obviously share a lot as well.

$$SWB_{ij} = a_j + b_j X_{ij} + \varepsilon_{ij}. \quad (5.1)$$

SWB_{ij} is the dependent variable and stands for the level of SWB (happiness or satisfaction) of individual i belonging to country j, a_j is a constant, X_{ij} denotes an individual determinant of SWB such as occupational status or a set of dummies in case of a categorical independent variable (for instance the individual's marital status), and ε_{ij} is an error term. Of only interest is the coefficient b_j—we are not interested in national differences in *levels* of SWB, which are captured by the intercept a_j. This coefficient may vary across nations (j), both in direction (positive/negative) and in size, which reflects the strength of the association between SWB and the independent variable of interest. The coefficients for all nations together capture the country variation in experienced preferences that provides the empirical input for the second step. We use a simple un-pooled approach to estimate this general regression model. This way, we obtain regression coefficients for every country, all individual determinants, and for both measures of SWB (happiness and satisfaction). Aknin et al. (2010) have used the same approach in their assessment of experienced preferences for prosociality in 136 countries. In our robustness checks, we also include results obtained using a Bayesian multilevel approach (cf. Chapter 3).

The second empirical step obtains our main result, evidence on the extent to which heterogeneity in experienced preferences is consistent, i.e. replicates itself across different measures of SWB. It involves a further examination of the regression results, specifically assessing the extent to which the country pattern in the estimated coefficients (the b's from Equation 5.1) is the same for the model with self-reported happiness and that with self-reported satisfaction with life as the dependent variable. We calculate correlations (r) between the happiness coefficients for each country and the satisfaction coefficients for each country, separately for the different individual determinants of SWB. These correlations represent a linear relationship between the two types of coefficients and show the extent to which country variation in the determinants of happiness (satisfaction) can be predicted from country variation in the determinants of satisfaction (happiness). The stronger the correlation between the estimated coefficients, the more consistent is country variation in experienced preferences across the two indicators, and the more confidence we can have in the reliability and validity of measured preference heterogeneity. Since we relate happiness and satisfaction to a set of situational factors and practices, we are able to make multiple assessments of consistency, individually for each independent variable.

Description of the Data

We conduct two separate psychometric evaluations of the consistency of heterogeneity in experienced preferences across different measures of SWB, each using a different dataset. Our first assessment uses data from the European Social Survey (ESS; e.g. Jowell and the Central Co-ordinating Team, European Social Survey 2007; see Chapter 4). The variables in this survey cover a variety of demographic and social characteristics, and also respondents' level of SWB. The survey's website gives more information and has the dataset available for downloading.[51] Dropping missing observations on either our two measures of SWB or on the selected situational factors typically leaves us with data on roughly 120,000 individuals residing in about 24 different countries.[52] The exact numbers depend on the variables included in the regression analyses.

The ESS measures happiness and life satisfaction by asking people "Taking all things together, how happy would you say you are?", and "All things considered, how satisfied are you with your life as a whole nowadays?" Answers come on scales ranging from 0 (Extremely unhappy / Extremely dissatisfied) to 10 (Extremely happy / Extremely satisfied) (see the appendix to Chapter 4).

Next to these two dependent variables, the ESS includes data on several situational factors and practices known to be significant determinants of SWB. As mentioned, we select those variables that are also available in more or less similar form for our second study based on a different, more diverse dataset. The individual variables we end up with are of two types, (semi-)continuous measures and categorical measures. The (semi-)continuous measures are:

- Income scale, ranging from 1 to 12;
- Health status, ranging from 1 (Very poor) to 5 (Very good);
- Sex;
- whether people belong to a Religious denomination, yes or no;
- Occupational status, based on Ganzeboom and Treiman's (1996) sorting of the International Standard Classification of Occupations (ISCO-88) into the International Socio-Economic Index of Occupational Status (ISEI), which runs from 16 (e.g. Farmhand) to 90 (Judge); and

[51] See http://www.europeansocialsurvey.org.

[52] These countries are Austria, Belgium, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, Finland, France, United Kingdom, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovakia, and Ukraine.

- whether people are a Supervisor at work, yes or no.

The categorical measures are:

- Marital status (e.g. "Married" or "Widowed");
- the frequency of Religious attendance (e.g. "Never" or "Every day");
- Meeting socially (e.g. "Never" or "Once a week");
- amount of Television watching (e.g. "No time at all" or "More than 3 hours per day");
- Education level (e.g. "Incomplete primary" or "Second stage of tertiary"); and
- Main activity (e.g. "Paid work" or "Unemployed, looking for a job").

In case of a (semi-)continuous independent variable like income scale we, of course, estimate only two coefficients per country, one for the model with happiness as the dependent variable and one for the model with satisfaction as the dependent variable. We recode the categorical measures into dummies with value 1 if the individual belonged to a particular class (for instance "Divorced") and 0 otherwise. We then include the set of dummies in our separate country regressions, taking the first answer category as the base category. To illustrate, for marital status, a categorical variable with five answer categories, we estimate 46 country models (23 with happiness and 23 with satisfaction as dependent variable), and each model rendered four regression coefficients, one for each dummy (4 x 46 = 184 in total). We assess consistency by calculating correlations independently for the four dummies.

The first three columns of Table 5.2a give more information on the independent variables and our happiness and satisfaction measures. These details include the number of individual observations (first column), means for the entire sample and standard deviations (second column), and the number of nations for which data are available (third column). For the dummies created from categorical independent variables, the mean value represents a percentage.

Our second assessment uses data from the long-running World Values Survey (WVS), which we also used in Chapter 3. Data come from all waves, more than 350,000 individuals in total (European Values Study Group and World Values Survey Association 2006; World Values Survey Association 2009). We refer to the description in Chapter 3 for more information about this dataset. We drop missing observations either on our measures of SWB or on the selected situational factors. The availability of the individual determinants of SWB varies a lot, but often times our sample comprised over 300,000 individuals from more than

95 societies.[53] Compared to the ESS sample we use for our first check of the psychometric quality of EP measures of preference heterogeneity, the WVS sample has much broader coverage and is less culturally homogeneous, including societies from all over the world, not just European countries.

The WVS measures happiness on a 4-point scale: "Taking all things together, would you say you are: 1-Not at all happy, 2-Not very happy, 3-Fairly happy, 4-Very happy." The life satisfaction item resembles more closely the one included in the ESS: "All things considered, how satisfied are you with your life as a whole these days?", with the answer scale ranging from "1-Dissatisfied" to "10-Satisfied" (see the second section and the appendix of Chapter 4). Our selected independent variables are almost exactly the same as in the first study. Dissimilarities are that income scale runs from 1 to 10, and that specific answer categories for some of the categorical measures do not fully overlap with the corresponding measures from the ESS. The most prominent differences concern the ESS items on "Meeting socially" and "Main activity," the WVS counterparts of which ask respondents about "Spending time with friends" and their "Employment status". The first three columns of Table 5.2b give more information on these variables, including specific answer categories for the categorical measures.

5.3.2. Basic Empirical Results

Assessment 1: Data from the European Social Survey (ESS)

Table 5.2a (last three columns) gives the results for our first psychometric evaluation of measured heterogeneity in experienced preferences. In order to avoid cluttering the text we

[53] The societies included in our empirical analysis of the WVS data are: Albania, Algeria, Andorra, Azerbaijan, Argentina, Australia, Austria, Bangladesh, Armenia, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Belarus, Canada, Chile, China, Taiwan, Colombia, Croatia, Cyprus, Czech Republic, Denmark, Dominican Republic, El Salvador, Ethiopia, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Japan, Jordan, South Korea, Kyrgyzstan, Latvia, Lithuania, Luxembourg, Malaysia, Mali, Malta, Mexico, Moldova, Morocco, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Puerto Rico, Romania, Russian Federation, Rwanda, Saudi Arabia, Singapore, Slovakia, Viet Nam, Slovenia, South Africa, Zimbabwe, Spain, Sweden, Switzerland, Thailand, Trinidad and Tobago, Turkey, Uganda, Ukraine, Macedonia, Egypt, Great Britain, Tanzania, United States, Burkina Faso, Uruguay, Venezuela, Serbia and Montenegro, Zambia, West Germany, Northern Ireland, and Serbia (98 in total), whereby we note that some of these no longer exist as independent and/or separate political entities.

do not report full quantitative results per country. We find substantial variation in the happiness and satisfaction coefficients (b's in Equation 5.1) across the 24 countries or so in our sample. This holds for both the (semi-)continuous determinants and the variables that we broke down into separate dummies. Although the sample includes European nations only, the relation between happiness/satisfaction and such factors as income scale, marital status, or religious attendance differs considerably between these societies. The variation in estimated happiness and satisfaction coefficients can be quite substantial as shown by the size of the standard deviations. For some measures or dummies the standard deviations are higher than the means of these regression coefficients. For other variables, notably income scale, health and occupational status, the size of the estimated coefficients is more dependable—always positive, as we would expect—but nonetheless varies greatly between nations.

Table 5.2a: Descriptive Statistics and Results for European Social Survey (ESS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 0-10 [123,929]	7.27 (1.98)	24	-	-	-
Life satisfaction, 0-10 [123,959]	6.94 (2.28)	24	-	-	-
Income scale, 1-12 [85,079]	6.37 (2.50)	22	.179 (.096)	.206 (.102)	.93
Health, 1-5 [124,355]	3.77 (.924)	24	.667 (.176)	.693 (.134)	.90
Sex, 1 = male [124,286]	.465 (.499)	24	.011 (.150)	-.008 (.143)	.92
Religious denomination, 1 = yes [121,373]	.621 (.485)	24	.110 (.173)	.161 (.209)	.84
Occupational status [109,029]	.420 (.165)	24	1.23 (.854)	1.45 (.870)	.91
Supervisor, 1 = yes [110,921]	.303 (.459)	24	.284 (.161)	.308 (.141)	.86
Legal marital status [82,716]					
Married	.536 (.499)	23	-	-	-
Separated	.016 (.125)	23	-1.05 (.407)	-.965 (.598)	.71
Divorced	.069 (.253)	23	-.848 (.263)	-.817 (.323)	.80
Widowed	.095 (.294)	23	-.912 (.367)	-.523 (.360)	.71

Table 5.2a, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Never married	.284 (.451)	23	-.163 (.302)	-.045 (.451)	.97
Religious attendance [123,951]					
Every day	.008 (.087)	24	-	-	-
More than once a week	.031 (.173)	24	.038 (.635)	.035 (.517)	.80
Once a week	.131 (.337)	24	-.018 (.599)	-.053 (.564)	.77
At least once a month	.106 (.307)	24	-.059 (.602)	-.098 (.604)	.77
Only on special holy days	.198 (.398)	24	-.078 (.599)	-.189 (.646)	.83
Less often	.203 (.402)	24	-.220 (.597)	-.329 (.609)	.81
Never	.324 (.468)	24	-.268 (.557)	-.414 (.610)	.74
Meet socially [124,156]					
Never	.022 (.146)	24	-	-	-
Less than once a month	.074 (.262)	24	.878 (.588)	.852 (.587)	.84
Once a month	.086 (.280)	24	1.31 (.660)	1.23 (.592)	.83
Several times a month	.181 (.385)	24	1.64 (.686)	1.60 (.606)	.87
Once a week	.180 (.384)	24	1.67 (.672)	1.63 (.598)	.85
Several times a week	.285 (.451)	24	1.85 (.695)	1.80 (.598)	.84
Every day	.172 (.377)	24	2.00 (.706)	1.91 (.605)	.82
Television watching [124,217]					
No time at all	.031 (.172)	24	-	-	-
Less than 0.5 hour	.057 (.233)	24	.331 (.383)	.269 (.353)	.83
Between 0.5 and 1 hour	.150 (.357)	24	.322 (.401)	.329 (.390)	.85
1 hour to 1.5 hour	.143 (.350)	24	.291 (.380)	.301 (.389)	.89

Table 5.2a, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
1.5 hour to 2 hour	.171 (.376)	24	.279 (.403)	.278 (.406)	.91
2 to 2.5 hour	.129 (.335)	24	.243 (.406)	.255 (.371)	.93
2.5 to 3 hour	.119 (.324)	24	.194 (.404)	.218 (.398)	.90
More than 3 hours per day	.200 (.400)	24	-.092 (.433)	.004 (.407)	.92
Education level [119,714]					
Incomplete primary	.039 (.194)	24	-	-	-
Primary or first stage of basic	.147 (.354)	23	.283 (.770)	.246 (.820)	.91
Lower secondary or second stage of basic	.219 (.414)	24	.452 (.783)	.329 (.898)	.91
Lower secondary or second stage of basic	.219 (.414)	24	.452 (.783)	.329 (.898)	.91
Upper secondary	.327 (.469)	23	.584 (.874)	.460 (.950)	.93
Post secondary, non-tertiary	.070 (.255)	21	.639 (.904)	.612 (1.02)	.90
First stage of tertiary	.155 (.362)	24	.760 (.925)	.705 (1.00)	.93
Second stage of tertiary	.042 (.200)	24	.887 (1.05)	.781 (1.10)	.92
Main activity last seven days [123,517]					
Paid work	.495 (.500)	24	-	-	-
Education	.088 (.283)	24	.350 (.335)	.492 (.468)	.97
Unemployed, looking for job	.032 (.176)	24	-.975 (.390)	-1.44 (.478)	.86
Unemployed, not looking for job	.015 (.123)	24	-1.02 (.448)	-1.23 (.528)	.64
Permanently sick or disabled	.022 (.148)	24	-1.53 (.494)	-1.72 (.413)	.73
Retired	.227 (.419)	24	-.354 (.415)	-.192 (.338)	.97
Community or military service	.002 (.039)	20	-.155 (.863)	.268 (.828)	.62

Table 5.2a, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Community or military service	.002 (.039)	20	-.155 (.863)	.268 (.828)	.62
Housework, looking after children, others	.106 (.308)	24	-.005 (.320)	-.041 (.328)	.95
Other	.012 (.108)	24	-.332 (.390)	-.332 (.427)	.79

Notes: Number of individual observations in square brackets. Standard deviations in parentheses. Table reports percentage responses per answer category for categorical independent variables. Individual and country observations are based on respondents with non-missing data on at least one of the SWB measures. Occupational status (ISEI) is divided by 100.

Comparing the happiness and satisfaction coefficients for the different nations (typically 24), we find considerable consistency across the two types of self-assessed well-being. This is evidenced by the high correlations (r) between the country estimates for the coefficients for the two indicators. These range from about 0.70 to more than 0.90. The country pattern in the estimated happiness (satisfaction) coefficients appears a powerful predictor of the country pattern in estimated satisfaction (happiness) coefficients. And this holds across the entire range of individual determinants of SWB. Hence, while there is substantial country variation in the relation between happiness/satisfaction and their determinants, this variation is actually quite consistent, and appears highly similar for the emotional evaluation (happiness) and the cognitive evaluation (satisfaction). This, in turn, indicates that heterogeneity in experienced preferences is of a structural rather than a random nature.

Assessment 2: Data from the World Values Survey (WVS)

The second assessment replicates and extends the first, using WVS data. Table 5.2b (last three columns) gives the results for this second psychometric evaluation of using SWB to measure preference heterogeneity. These are analogous to those obtained in the first assessment. The analysis reveals substantial heterogeneity across countries in the relation between happiness and satisfaction and their determinants. In fact, overall country variation in the structure of SWB appears higher than in the previous study. This is what we would expect

given that the first assessment comprises only 24 countries, all from Europe, whilst the sample for Assessment 2 covers more nations that are likely less culturally homogeneous and entertain generally more diverging preferences.

Concerning consistency, a divergence between the first and the second assessment is also readily apparent. Bottom line, the extent to which variation in the structure of happiness replicates itself in structural variation in satisfaction is a good deal lower in our second assessment than it is in our first. Most strikingly, with dummies indicating individuals' marital status as independent variables, the correlations reach as low as 0.17 and only as high as 0.77—compared with r's between 0.62 and 0.97 in the previous study. Another remarkable difference is the effect of health on SWB, for which consistency drops considerably, from r = 0.90 to r = 0.49.

Table 5.2b: Descriptive Statistics and Results for World Values Survey (WVS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [344,363]	3.03 (.740)	98	-	-	-
Life satisfaction, 1-10 [349,518]	6.65 (2.45)	98	-	-	-
Income scale, 1-10 [307,118]	4.65 2.44)	98	.049 (.030)	.204 (.130)	.76
Health, 1-5 [302,669]	3.78 (.910)	94	.259 (.081)	.689 (.184)	.49
Sex, 1 = male [349,955]	.48 (.500)	98	-.011 (.069)	-.025 (.192)	.78
Religious denomination, 1 = yes [260,295]	.802 (.398)	77	.048 (.113)	.156 (.435)	.65
Occupational status [28,745]	.426 (.164)	30	.415 (.226)	1.59 (.818)	.66
Supervisor, 1 = yes [68,637]	.309 (.462)	69	.094 (.090)	.367 (.279)	.73
Marital status [349,667]					
Married	.588 (.492)	98	-	-	-
Living together as married	.052 (.221)	91	-.055 (.189)	-.137 (.557)	.17
Divorced	.035 (.185)	96	-.287 (.196)	-.681 (.490)	.30
Separated	.016 (.126)	96	-.326 (.228)	-.859 (.687)	.54

Table 5.2b, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Widowed	.065 (.247)	97	-.284 (.150)	-.541 (.456)	.70
Single/Never married	.225 (.418)	98	-.054 (.099)	-.098 (.371)	.77
Religious attendance [333,097]					
More than once a week	.123 (.329)	96	-	-	-
Once a week	.184 (.388)	96	-.079 (.126)	-.195 (.419)	.62
Once a month	.114 (.318)	96	-.097 (.121)	-.284 (.459)	.45
Only on special holy days	.146 (.353)	96	-.093 (.130)	-.276 (.429)	.59
Other specific holy days	.021 (.143)	46	-.101 (.273)	-.306 (.686)	.77
Once a year	.072 (.259)	96	-.106 (.185)	-.358 (.571)	.62
Less often	.102 (.303)	95	-.135 (.143)	-.414 (.517)	.59
Never/practically never	.238 (.426)	95	-.162 (.167)	-.486 (.510)	.63
Spend time with friends [101,158]					
Weekly	.521 (.500)	71	-	-	-
Once or twice a month	.297 (.457)	70	-.043 (.060)	-.175 (.176)	.41
Only a few times a year	.130 (.336)	70	-.122 (.099)	-.423 (.407)	.51
Not at all	.052 (.223)	69	-.269 (.191)	-.814 (.746)	.63
Television watching [117,092]					
Do not watch TV	.104 (.305)	72	-	-	-
1-2 Hours per day	.423 (.494)	72	.124 (.180)	.338 (.519)	.75
2-3 Hours per day	.253 (.435)	71	.140 (.178)	.336 (.558)	.71
More than 3 hours per day	.220 (.414)	71	.127 (.213)	.256 (.586)	.76
Education level [269,722]					

Table 5.2b, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Incomplete elementary	.131 (.338)	97	-	-	-
Completed elementary	.151 (.358)	97	.065 (.175)	.209 (.479)	.65
Incomplete secondary: technical / vocational	.081 (.272)	92	.123 (.234)	.332 (.931)	.80
Complete secondary: technical / vocational	.164 (.370)	96	.182 (.185)	.332 (.931)	.71
Incomplete secondary: university-preparatory type	.094 (.292)	95	.159 (.187)	.449 (.594)	.64
Complete secondary: university-preparatory	.168 (.374)	95	.204 (.189)	.619 (.580)	.67
Some university without degree / Lower-level tertiary	.071 (.258)	96	.229 (.210)	.746 (.680)	.68
University with degree / Upper-level tertiary	.139 (.346)	95	.249 (.212)	.840 (.657)	.73
Employment status [342,636]					
Full time	.377 (.485)	98	-	-	-
Part time	.074 (.262)	96	-.015 (.098)	-.112 (.318)	.47
Self employed	.094 (.292)	98	-.021 (.089)	-.022 (.398)	.59
Retired	.135 (.342)	98	-.110 (.152)	-.208 (.410)	.77
Housewife	.145 (.352)	98	.010 (.120)	-.084 (.453)	.73
Students	.072 (.259)	98	.080 (.113)	.227 (.462)	.71
Unemployed	.084 (.277)	97	-.179 (.134)	-.760 (.453)	.72

Table 5.2b, Continued.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Other	.019 (.135)	87	-.048 (.238)	-.362 (.787)	.54

Notes: See Table 5.2a.

An explanation for these differences in consistency between the ESS sample and the WVS sample may be that the scales for the two types of well-being assessments in the WVS are not precisely matched, with happiness ranging from 1 to 4, and satisfaction with life from 1 to 10, whilst they were equivalent in the first study. Most obvious implication of this scale incongruence is that the size of the happiness and satisfaction coefficients cannot be put side by side. In addition, this inexact matching likely results in added random noise in the estimated happiness coefficients relative to the estimated satisfaction coefficients. This may have lowered the correlations between the two types of coefficients.

We expand on these and other features of our two assessments and how they may affect our findings on the consistency of country variation in experienced preferences next.

5.4. ROBUSTNESS OF THE PSYCHOMETRIC QUALITY OF HETEROGENEOUS EXPERIENCED PREFERENCES

Concerning the question whether country heterogeneity in experienced preferences is consistent across the two types of well-being assessments, we find that it tends to be, but that consistency is a function of different characteristics of the analyses. These are, in no particular order:

1. the choice of the base category for dummified categorical variables,
2. the number of observations (per independent variable),
3. the type of coefficients for the independent variables we estimate in our separate country regression, standardized or unstandardized,
4. the matching of the scales used to measure happiness and satisfaction respectively, specifically whether they span the same range of possible answers (e.g. from 0 to 10), and

5. the specific statistical approach used, especially with regards to the way the variation in the sample is modeled, separately for each country or using a more advanced statistical technique namely multilevel or hierarchical linear modeling.

The discussion below considers these specific features and how they may affect our findings. By performing additional tests that take these specifics of the assessments into account we show the robustness of our original findings—or lack thereof.

5.4.1. Base Category

A first feature is the base category chosen for the categorical independent variables. Table 5.3 illustrates the influence of base category for the case of religious attendance using the ESS sample (seven answer categories; six dummies).

Table 5.3: Robustness to Base Category of Correlations between Estimated Happiness and Satisfaction Coefficients (ESS).

Variable	Empirical results	
	Correlation (r) between estimated happiness and satisfaction coefficients, original	Correlation (r) between estimated happiness and satisfaction coefficients, new
Religious attendance		
Every day	-	.74
More than once a week	.80	.82
Once a week	.77	.81
At least once a month	.77	.83
Only on special holy days	.83	.80
Less often	.81	.70
Never	.74	-

Notes: See Table 5.2a.

Naturally, changing the base category from "Every day" (0.8% of all respondents) to "Never" (32.4% of all respondents) affects the correlation coefficients showing the consistency of cultural variation in the structure of SWB. Correlations for some categories increase somewhat, whilst for others they decrease a bit (with the original correlation for "Never", r = 0.74, now found for "Every day"). A priori we have no reason to assume that one base category is superior to another, in the sense of providing a better picture of heterogeneity in experienced preferences, though future research may investigate this issue further. In addition, the choice of base category may follow naturally from the specific variation in experienced preferences one seeks to investigate.

5.4.2. Number of Observations

Both county heterogeneity in experienced preferences and the consistency of this heterogeneity are liable to be affected by the number of observations per country (and per country per independent variable / answer category). Simply put, more data enhances the accuracy/precision of our estimates and findings. What is more, in some countries certain variables have only a minimum number of observations. As a result, for these countries happiness and satisfaction coefficients concerning these variables are estimated with limited accuracy/precision. This, in turn, affects measured preference heterogeneity and its consistency across SWB indicators.

Table 5.4a: Marital Status, Minimum 0.5% of Observations "Living together as married" (WVS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [301,693]	3.03 (.737)	84	-	-	-
Life satisfaction, 1-10 [303,811]	6.71 (2.41)	84	-	-	-
Marital status [307,756]					
Married	.577 (.494)	84	-	-	-
Living together as married	.059 (.235)	84	-.040 (.131)	-.179 (.422)	.64
Divorced	.038 (.192)	84	-.295 (.181)	-.663 (.446)	.37
Separated	.018 (.131)	84	-.324 (.225)	-.800 (.631)	.53
Widowed	.069 (.253)	84	-.290 (.156)	-.536 (.454)	.72
Single/Never married	.239 (.427)	84	-.057 (.100)	-.091 (.391)	.79

Notes: See Table 5.2b.

Most prominently, we expect this to be the case with marital status, for which the frequency of occurrence of certain states such as divorce or unmarried cohabitation are likely to be extremely low (even nil) in some countries, whilst it is much more common in others. This feature may account, at least partly, for the finding of rather low consistency for certain

marital status dummies in the second study. We investigate this possibility by re-analyzing heterogeneity in the relation between marital status and happiness/satisfaction in the WVS sample and selecting only countries with a minimum number of observations (> 0.5%) on the dummies indicating whether people were "Living together as married" or "Divorced". Tables 5.4a and 5.4b give the results, which compare to the results depicted in Table 5.2b.

Table 5.4b: Marital Status, Minimum 0.5% of Observations "Divorced" (WVS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [313,679]	3.01 (.739)	89	-	-	-
Life satisfaction, 1-10 [315,942]	6.61 (2.45)	89	-	-	-
Marital status [319,799]					
Married	.584 (.493)	89	-	-	-
Living together as married	.050 (.219	84	-.057 (.195)	-.179 (.446)	.28
Divorced	.038 (.192)	89	-.303 (.182)	-.666 (.459)	.46
Separated	.016 (.126)	89	-.330 (.230)	-.853 (.649)	.51
Widowed	.068 (.252)	89	-.283 (.154)	-.538 (.451)	.71
Single/Never married	.242 (.428)	89	-.058 (.099)	-.100 (.378)	.77

Notes: See Table 5.2b.

The effect of dropping only a few countries that do not meet our percentage criterion is rather strong. On the one hand, results still demonstrate substantial country differences in the relationship between marital status and SWB. On the other hand, the consistency across types of self-assessed well-being as reflected by the correlations between happiness and satisfaction coefficients is greatly increased.

We can further apply a minimum percentage of observations criterion to the measure of health status in the WVS sample. This allows us to delve more deeply into the role the number of observations plays in consistency of measured preference heterogeneity across indicators of SWB. As it turns out, some countries lack observations, meaning that not the entire range of health ratings, 1-2-3-4-5, is present. The twin Tables 5.5a and 5.5b display the

results for experiments with two different criteria, one demanding a minimum of 0.5% of observations per score, and one demanding 0.05% (cf. Table 5.2b).

Table 5.5a: Health Status, Minimum 0.5% of Observations on Each Rating (1-2-3-4-5) (WVS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [203,856]	2.98 (.747)	46	-	-	-
Life satisfaction, 1-10 [206,894]	6.60 (2.45)	46	-	-	-
Health, 1-5 [180,885]	3.73 (.946)	46	.253 (.061)	.698 (.152)	.69

Notes: See Table 5.2b.

Again, we find that the number of observations plays a large role. Consistency of heterogeneity in health's SWB effect was much lower in the second of our two assessments than in the first. Requiring a minimum number of observations per health score leads to a much higher correlation between the happiness and the satisfaction coefficient concerning health also for the WVS sample. Implication is that the study of heterogeneous experienced preferences will benefit from ensuring that enough observations are available.

Table 5.5b: Health Status, Minimum 0.05% of Observations on Each Rating (1-2-3-4-5) (WVS).

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [290,473]	3.03 (.739)	69	-	-	-
Life satisfaction, 1-10 [295,371]	6.75 (2.42)	69	-	-	-
Health, 1-5 [253,406]	3.76 (.921)	69	.249 (.057)	.679 (.143)	.65

Notes: See Table 5.2b.

5.4.3. Unstandardized versus Standardized Coefficients

Thus far, our analyses involved unstandardized happiness and satisfaction coefficients (b's). Some may want to argue, however, that it is more appropriate to look at standardized, so-called beta coefficients instead. Standardized coefficients—taking all variables, independent and dependent, subtracting the mean and dividing by the standard deviation—have no dimension. Rather, after standardization the estimated regression coefficient represents how many standard deviations the dependent variable (happiness or satisfaction) changes as a result of a one standard deviation change in an independent variable (e.g. occupational status). In principle, using standardized instead of unstandardized coefficients may affect the consistency of measured preference heterogeneity across SWB indicators. To investigate this issue, we rerun the regressions included in Table 5.2a (ESS). As it does not make sense to standardize dummy coefficients, we limit attention to the continuous independent variables, income scale, health, and occupational status. Table 5.6 gives the results.

Table 5.6: Robustness of Correlations: Standardized versus Unstandardized Coefficients (ESS).

	Empirical results	
Variable	Correlation (r) between *unstandardized* happiness and satisfaction coefficients	Correlation (r) between *standardized* happiness and satisfaction coefficients
Income scale, 1-12	.93	.84
Health, 1-5	.90	.78
Occupational status	.91	.90

Notes: See Table 5.2a.

As it turns out, consistency is somewhat lower when we use standardized instead of unstandardized coefficients. Apparently, switching to standardized coefficients removes a small part of the group variation in experienced preferences that is structural, i.e. that is independent of the particular indicator of SWB used as dependent variable in the separate country regressions. A reason may be that countries' average level of self-reported happiness or satisfaction affects the size of unstandardized coefficients. This is a possibility, even though our regressions included intercepts. Moreover, it can account for the fact that there appear to be country-specific differences in unstandardized happiness and satisfaction coefficients that do not exist in standardized coefficients. Notwithstanding, EP measures show high levels of consistency, irrespective of the specific approach used to estimate heterogeneous experienced utility functions.

5.4.4. Inexactly Matched Subjective Well-Being Scales

Scale incongruence is another factor that likely affects the correlation between the happiness and satisfaction coefficients. We assess this potentiality by reconsidering the consistency of variation in the determinants of SWB across the two types of self-assessed well-being for the 24 countries included in the ESS but using data from the WVS. Because we keep the same set of countries, deviations in consistency cannot be attributed to deviations in the country composition of the dataset. This means that any differences in consistency we may find between the first assessment for which we used ESS data and these new results must be due to other factors, notably the matching of the happiness and satisfaction measures. The WVS includes corresponding data for most measures and countries in the ESS sample, but we only report quantitative results for the continuous variables (as for Table 5.6). Results are presented in Table 5.7.

Table 5.7: Descriptive Statistics and Results for ESS Countries in the WVS.

Variable names and descriptives			Empirical results		
Variable	Individual mean	Number of societies	Mean estimated happiness coefficients	Mean estimated satisfaction coefficients	Correlation (r) between regression coefficients
Happiness, 1-4 [108,282]	3.06 (.673)	24	-	-	-
Life satisfaction, 1-10 [110,508]	7.08 (2.17)	24	-	-	-
Income scale, 1-10 [88,851]	4.87 (2.49)	24	.044 (.022)	.144 (.098)	.93
Health, 1-5 [81,577]	3.74 (.945)	22	.219 (.045)	.653 (.128)	.84
Occupational status [20,088]	.428 (.161)	21	.373 (.207)	1.45 (.672)	.76

Notes: See Table 5.2a.

Correlations are lower than in the original study (cf. Table 5.2a). This suggests that scale mismatch indeed decreases the consistency of heterogeneity in experienced preferences measured across different SWB indicators. It therefore seems wise to discount the relatively low levels of consistency found in our second assessment of the psychometric quality of heterogeneous experienced preferences (see Table 5.2b) as they appear partly driven by the lack of correspondence between the happiness and satisfaction measures included in the

WVS.[54] Put differently, the results of our replication of Assessment 1 depicted in Table 5.7 strengthen the conclusion that country variation in experienced preferences may replicate itself across measures of SWB to a considerable extent. In addition, future work studying the consistency of systematic variation in experienced preferences may aim to make the different SWB indicators used as congruent as possible.

5.4.5. Bayesian Multilevel Results

Our statistical approach has been deliberately simple. We may wonder, however, how the use of more advanced techniques would affect consistency. Our data fit the mold for multilevel or hierarchical linear modeling very well, with individual observations nested in higher level units (i.e. countries). It therefore seems logical to examine the see how applying this technique affects our results. A feature of multilevel modeling, which we already applied in Chapter 3, is that it is Bayesian and that the separate country happiness and satisfaction coefficients are posterior estimates, which are shrunk towards the sample average for these coefficients, this way borrowing strength from information obtained from all available data (Gelman and Hill 2007; Raudenbush and Bryk 2002; Snijders and Bosker 1999).

As it turns out, using multilevel modeling and getting Bayesian posteriors for our two types of coefficients has almost no effect; whatever little effect there is, is to increase consistency, which is not inconsistent with the idea of shrinkage towards the mean slope coefficient of the sample and borrowing strength (Table 5.8).

Table 5.8: Robustness of Correlations: Posterior Multilevel versus Traditional Coefficients (ESS).

	Empirical results	
Variable	Correlation (r) between *posterior multilevel* happiness and satisfaction coefficients	Correlation (r) between *traditional* happiness and satisfaction coefficients
Income scale, 1-12	.93	.93
Health, 1-5	.91	.90
Occupational status	.92	.91

Notes: See Table 5.2a.

[54] Of course, this is not the only factor that accounts for the lower level of consistency. Another factor that likely is important is that using WVS data, we have a lower number of observations compared to the ESS data (see above).

This concludes our assessment of the robustness of the consistency of EP measures of preference heterogeneity across different types of well-being assessments (cognitive and affective). The next section provides a summary of the findings of this chapter and a conclusion.

5.5. CONCLUDING AND SUMMARIZING REMARKS

As discussed at length in the previous chapter, SWB is a survey-based measure obtained by simply asking people to answer certain questions. It therefore is susceptible to the problems survey data can typically have (see Section 2.4). For SWB this appears to hold even more than for most other variables measured using questionnaires. After all, it is not evident what one is measuring when asking individuals how happy or how satisfied they are and then supplying them with a scale ranging from 1 to 10 to record their answer on. At the core, this question simply asks people to put a label on their life. There is no rationale for asking this question with these specific response categories other than the expectation that the resulting measure will be meaningful.

Considering all this, it might be surprising to find that measures of SWB are actually reasonably reliable and valid, i.e. able to capture something meaningful. As the review in this chapter shows, indicators of SWB do provide important and useful information on how well people or societies as a whole are doing. Even if the measured level of SWB indeed is just a simple label people put on their life when asked upon, it appears to matter a lot. For instance, it correlates with many factors generally considered important for well-being, such as health, income, and good relationships. In addition, it has physiological counterparts, notably indicators of SWB correlate with patterns of brain activity and blood pressure levels.

We argue that for many of its uses, the reliability and validity of SWB indicators implies that we can be confident about the results obtained by applying them. In our view, the more indirect the use of SWB, the less the psychometric quality of the indicators suffices as a guarantee that their application renders meaningful results. This would be the case for SWB-based valuation exercises such as the ones discussed in the section on policy uses of SWB in Chapter 4, but, more importantly, also for the estimation of heterogeneous experienced utility functions to measure preference heterogeneity. Solace may be found in additional psychometric evaluation; and this is what we have done for EP measures of preference heterogeneity.

Our approach to assessing the psychometric quality of EP measures of preferences and heterogeneity therein has been to look at the extent to which country variation in experienced preferences is consistent across two different types of well-being assessments, namely self-reported happiness and satisfaction with life. Of these, the former is associated more with the affective part of SWB and the latter more with the cognitive part of SWB. Our comprehensive analyses included a diverse set of individual determinants of SWB and up to 98 nations. Country-specific regression models rendered estimated happiness coefficients and satisfaction coefficients, the correlations between which demonstrate consistency. Results show marked consistency with correlations between the two types of coefficients varying between roughly 0.60 and 0.90, reaching up to 0.95. The cross-country pattern of heterogeneity in the determinants of SWB thus replicates itself in the cross-country pattern of heterogeneity in the determinants of satisfaction with life to a considerable degree. The high level of consistency indicates that group variation in experienced preferences is mostly of a structural nature as opposed to being random. As such, we deem EP measures reliable and valid. Robustness checks of our initial results confirm this conclusion. Additional examination further reveals that increasing the number of observations substantially raises consistency of measured heterogeneity in experienced preferences across different indicators of SWB.

Table 5.9: A Summary Evaluation of the Usefulness of the EP Approach to Measuring Heterogeneous Preferences and Studying the Economics of Preference Heterogeneity.

	New evaluation scores	Old evaluation scores
Psychometric quality	●●●●	● - ●●●●
Usability		
Availability	●●●●	●●●●
Scope	●●●●	●●●●
Resolution	●	●
Explanatory capacity	●●●●●	●●●●●

Notes: See Table 2.1.

We can summarize our findings by updating the prior assessment of the psychometric quality of the EP approach and EP measures of preference heterogeneity depicted in Table 2.1. Table 5.9 presents our revised psychometric evaluation. Take-home message is that the assessment presented in this chapter has taken away much of the uncertainty that previously surrounded the EP approach's ability to render meaningful preference measures. What is more, the level of consistency found assures us of the potential the EP method has in the study of the economics of preference heterogeneity. We turn to applying the EP method for this purpose in the third part of this thesis, Chapters 6 and 7.

Summary and Conclusion of Part II

The two chapters in the second part of this thesis have taken up three key issues concerning subjective well-being (SWB) and the measurement of (heterogeneous) experienced preferences. Firstly, this part has introduced the empirical SWB construct and measures of SWB, which, though much-researched in psychology, are relatively new to economics. We have elaborated the definition of SWB as "a broad category of phenomena that includes people's emotional responses, domain satisfactions, and global judgments of life satisfaction" (Diener et al. 1999: 277), and have given key examples of measurement instruments and SWB scales. The latter vary from a single question simply asking people how happy they are or how satisfied with their life as a whole to detailed questionnaire items about a variety of different emotions and their frequency of occurrence during specific episodes. Similar differences in sophistication can be found in the methods by which these scales are implemented: the most ubiquitous SWB data come from one-time cross-country surveys but it is also possible to track individuals for prolonged periods and ask them to report on their SWB at random intervals during the day. Secondly, we have evaluated the reliability and validity of SWB indicators. Surveying existing work on the psychometric properties of SWB, we find that indicators of SWB are not without problems but are nonetheless reasonably reliable and valid. The reliability and validity of SWB indicators is a prerequisite for their use in the measurement of experienced preferences and heterogeneity therein.

The main contribution of Part II is its empirical assessment of the psychometric quality of EP measures of preference heterogeneity. Part I found, partly by elimination, that of the three approaches to measuring preferences and assessing preference heterogeneity, the EP approach has the most promise for the study of the economics of preference heterogeneity. This conclusion was preliminary, however, as this potential remained largely unproven. Particularly absent in the literature applying the EP approach to the study of the economics of heterogeneous preferences was evidence on the reliability and validity of the measures thus obtained. Our empirical analysis has demonstrated the psychometric quality of EP measures thus taking away some of the doubt that may have surrounded the only recently developed approach of estimating experienced utility functions to measure preferences and heterogeneity therein.

We summarize the findings of Part II of the thesis, specifically Chapter 5, in the form of an update to Table I, which contains a summary of chief results of the first part of the

thesis. Table II below depicts the updated comparative evaluation of the three approaches—RP, SP and EP.

Table II: Summary of the Comparative Evaluation of the RP, SP and EP Approaches and Measures.

Approach	Revealed Preferences	Stated Preferences	Experienced Preferences
Psychometric quality	● - ●●●●●	●	●●●●
Usability			
Availability	●	●●●●●	●●●●
Scope	●	●●●●●	●●●●
Resolution	●●●●●	●●●●●	●
Explanatory capacity	●	●	●●●●●

Notes: See Table I and Table 5.9.

Since our scrutiny of the psychometrics of both SWB and EP measures of heterogeneous preferences has confirmed the potential of the EP approach that became apparent in the preceding part, we now have a solid foundation for applying the EP method to study the economics of preference heterogeneity. The third and final part of this thesis consists of two such applications whereby the EP approach is used to shed new light on some long-standing issues in the economics of preference heterogeneity.

Part III: Assessments of Heterogeneity in Experienced Preferences

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of *science*, whatever the matter may be.

- Lord Kelvin (1883)•

• This is the quote exactly as it appeared in Thomson (1889: 73-74); see Merton et al. (1984).

Chapter 6

Does a Protestant Work Ethic Exist? Evidence from Religious Variation in the Psychic Cost of Unemployment

6.1. INTRODUCTION

More than a century after its publication, Max Weber's *The Protestant Ethic and the Spirit of Capitalism* (Weber 1904/1905 [1930]) continues to inspire social scientists in many disciplines. A large stream of work in the social sciences has built on the idea that religious values can explain social and economic developments. This work ranges from studies with a historical perspective (e.g. Huntington 1993, 1996; Landes 1998) to quantitative research assessing religious variation in value preferences, and also relating these to GDP growth, savings rates, democracy and other societal-level outcomes (e.g. Arruñada 2010; Barro and McCleary 2003; Guiso et al. 2003, 2006; Inglehart et al. 2002; McCleary and Barro 2006; Noland 2005; Rowley and Smith 2009; Schaltegger and Torgler 2010; Torgler and Schaltegger 2009; see, also, Chapter 3). Paradoxically, much of this research has failed to find support for the thesis of Weber that originally inspired the literature. Any convincing relation between Protestantism on the one hand and work ethic or economic development on the other has yet failed to materialize in the data (e.g. Lehmann and Roth 1993; Iannaccone 1998: 1474-1478; Delacroix and Nielsen 2001).[55] In fact, researchers have reported evidence that Protestants overall value work less than people from other religions do (Norris and Inglehart 2004; Weil 2009; see Furnham 1984, 1990 and Jones 1997 for surveys of this type of empirical studies of work values and Protestantism). Furthermore, various authors have found intriguing and counterintuitive negative associations between work-related practices and values retrieved from values surveys (e.g. Bertrand and Mullainathan 2001; McCleary and Barro 2006; Weil 2009), implying that even if Protestants were to have a stronger work ethic, such a work ethic does not translate into a higher rate of employment or into longer working hours.

In this chapter, we apply the EP approach to provide a comprehensive analysis of religious variation in work ethic, specifically the existence of the Protestant work ethic hypothesized by Weber. We find that much of the ambiguity and confusion in the existing literature may derive from a lack of valid measurement of the relevant preference for work. Indeed, the counterintuitive findings just mentioned and the generally mixed results on a specific Protestant work ethic, both of which derive from quantitative analyses of

[55] Iannaccone (1998: 1474) thus concludes that "[i]ronically, the most noteworthy feature of the Protestant Ethic thesis is its absence of empirical support." In so far as people have found support for the idea of a link between Protestantism and economic prosperity, it has been argued that the more plausible route runs via literacy levels rather than the prevalence of a "capitalist spirit," as Weber wants it (Becker and Woessmann 2009, 2010).

questionnaire items on the value people attach to work (relative to leisure), strongly resemble the kind of response patterns associated with the marginal preferences problem (see Chapter 3). Realizing that people tend to profess to value work (or leisure) highly typically because they have little of it, such counterintuitive and intriguing findings suddenly make a lot more sense.

In our approach we use the degree to which having a job adds to people's well-being (specifically their SWB) as an indicator of the importance people attach to work and, hence, of their work ethic. There is a large literature showing and theoretically arguing that unemployment is associated with large psychic costs, even when its economic consequences are controlled for. We introduce religious heterogeneity in the adverse well-being effect of being unemployed relative to having a full-time job. Allowing the psychic cost of unemployment to vary across individuals with different denomination gives us the direct means to assess how much religious background matters for the value people attach to work.

Following Weber, we hypothesize two possible effects of Protestantism, one operating at the individual level and one at the societal level. Firstly, the psychic cost of not having a job may be higher for Protestants than for individuals from other denominations. Secondly, employment may hurt individuals from Protestant societies more than those living in other societies. We examine a broad sample, comprising almost 150,000 individuals from 82 societies, and find strong evidence for Protestantism's influence on work ethic: both individual Protestants and Protestant societies appear to value work much more highly, where the societal effect appears to dominate. Several robustness checks show that the findings are not driven by other factors such as higher materialism amongst Protestants and/or in Protestant societies. Exploring possible channels we find that the effect of Protestantism does not run through social stigma effects either. Hence, more than a century after Weber coined his original thesis we find support for a specific Protestant work ethic.

The structure of this chapter is as follows. In the next section we discuss the Weber thesis in more detail, looking at contemporary interpretations and efforts to find empirical support for the thesis. Section 6.3 takes the literature on SWB and unemployment to propose an operationalization of work ethic in terms of experienced preferences. We then apply this operationalization to the problem of the relation between Protestantism and work ethic. Our empirical strategy and the results of this endeavor are discussed in Sections 6.4 and 6.5. We round off with a discussion and conclusion in Section 6.6.

6.2. RELIGION AND WORK ETHIC: WEBER'S THESIS

Weber's (1904/1905 [1930]) *The Protestant Ethic and the Spirit of Capitalism* goes into history as one of the most frequently cited books in social science. His argument about the relation between Protestantism and capitalism has spawned an extensive and diverse literature dealing with the effect of religious values on economic performance. Some of this literature has retained Weber's original focus on Protestantism. Other contributions have reworked the argument, applying it to other religions such as Catholicism (Tawney 1926) or, lately, to Confucianism (Harrison 1992; Harrison and Huntington 2000; Kahn 1979; Eungi Kim and Park 2003). Still others have quantitatively assessed the role of non-religiously specified sets of values (e.g. Granato et al. 1996). More recently, general associations between religion and economic outcomes have been analyzed empirically (Barro and McCleary 2003; Guiso et al. 2006; Noland 2005). Throughout this literature, Weber is commonly referred to as the person starting the debate about the link between religious ethics and economic growth.

6.2.1. The Weber Thesis Theoretically

Before we delve into empirical tests of the Weber thesis, it pays to briefly go back to the original argument of *The Protestant Ethic*. As Giddens (1992) and many others have emphasized, Weber's famous essay can be approached on many different levels. *The Protestant Ethic* establishes a historical relation between the emergence of capitalism as a dominant economic system in Western Europe and North America and the Protestant reformation centuries earlier. More in particular, Weber draws attention to the peculiar ascetic ethical system propagated in specific sub-denominations of Protestantism, namely Calvinism, Pietism, Methodism and Baptism. Here originated the idea of a "calling," a perception of one's work and other economic activities as God-given duties. The emphasis on worldly activity as a means to prove one's faith in these strands of Protestantism eventually evolved, through a process of rationalization, into what Weber calls the "spirit of capitalism;" the idea that working for the purpose of profit is a moral good in itself. Weber (1904/1905 [1930]: 19) writes:

> … one's duty in a calling, is what is most characteristic of the social ethic of capitalistic culture, and is in a sense the fundamental basis of it. It is an obligation which the individual is supposed to feel and does feel towards the content of his

> professional activity, no matter in what it consists, in particular no matter whether it appears on the surface as a utilization of his personal powers. Or only of his material possessions (as capital).

In contrast to the common interpretation of *The Protestant Ethic* (e.g. Becker and Woessmann 2009, 2010; Granato et al. 1996; Weil 2009), there is little in the original text to suggest that Weber saw a causal relation between being Protestant and enjoying economic prosperity. Rather than that, he was trying to explain the initial origin of modern industrial capitalism in Northwest Europe and North America, relating it to values that could be retraced historically to a specific religious ethics. Weber argues that modern capitalism's outstanding feature—compared to previous capitalist practices—is a set of values that is religious in origin, but has been rationalized and secularized over time.[56] What is more, a link with economic growth and prosperity is largely absent in the work.[57] Nevertheless, the "common interpretation" (Delacroix and Nielsen 2001) of *The Protestant Ethic*, seeing in Protestantism a cause of economic progress, has taken a life of its own.

A problem with this common interpretation is that it transforms Weber's argument into two controversial theses rather than one, rendering testing difficult. First, it argues that Protestantism results in a strong work ethic, and second, it claims that a strong work ethic is a main determinant of economic growth and prosperity. Elsewhere, it has been shown that the latter relation is not that straightforward. Some authors have reported empirical evidence that people in poor countries attach more importance to work than people in developed countries do (McCleary and Barro 2006; Weil 2009). Also, it has been pointed out that inhabitants of some of the most rapidly growing economies of the past century—for example Japan, South Korea, and Malaysia—have historically been viewed as lacking a strong work ethic (Alatas 1977; Chang 2007; Landes 1998). Often, arguments about any relation between a work ethic and economic prosperity have served more as a legitimization than an explanation of economic inequality (Alatas 1977; Said 1978).

[56] Weber (1904/1905 [1930]: 27) writes: "the cultural consequences of the Reformation were to a great extent, perhaps in the particular aspects with which we are dealing predominantly, unforeseen and even unwished-for results of the labors of the reformers. They were often far removed from or even in contradiction to all that they themselves thought to attain."

[57] The awareness of such a link may have been Weber's motivation to write *The Protestant Ethic,* however.

6.2.2. The Weber Thesis Empirically

The above caveats lead us to adopt a more limited focus here, dealing only with the relation between Protestantism and work ethic, without going into the effects on economic performance. An interesting recent empirical investigation into this relation is by Norris and Inglehart (2004). The idea in this study is that, if Weber's thesis is correct, Protestantism should "have left an enduring legacy in values that still remains visible today" (Norris and Inglehart 2004: 162). To test this idea they construct a multidimensional measure of work ethic using a selection of items from the World Values Survey (WVS). Specifically, their index combines questions asking about, amongst others, which aspects people find important in a job (e.g. "an opportunity to use initiative," "good hours," and "good pay"), and the extent to which people agree with the statement that "work is a duty towards society", and that "people who don't work turn lazy" (Norris and Inglehart 2004: 163). To their surprise, the analysis shows that people living in Protestant societies have a weaker work ethic than many individuals from other religious cultures. Norris and Inglehart (2004) conclude that the Weber thesis is to be dismissed.

A closer look at the recent work of Schaltegger and Torgler (2010) and Torgler and Schaltegger (2009), which we briefly discussed in Chapter 3, confirms the negative association between Protestantism and work ethic. On the one hand, Torgler and Schaltegger's (2009) cross-country analysis finds that the link between religious belief in hell and work ethic—reported earlier by McCleary and Barro (2006)—is statistically significantly stronger the higher the share of Protestants in society. On the other hand, the association is also positive the higher the share of Catholics in a country though this relation is about five times weaker and not statistically significant. More importantly, the direct effect of share of Protestants on work ethic is, in fact, statistically significantly negative and so is the effect of per-capita GDP—the latter confirming the findings by McCleary and Barro (2006) and Weil (2009).[58] In similar fashion, Schaltegger and Torgler's (2010) panel study of religious variation in value attached to work for a sample of 17,000 individuals from 16 European countries establishes a channel linking Protestant denomination to work ethic through education: for Protestants, a higher education level is associated with stronger agreement with the statement that work should come first even if it means less spare time (cf. Chapters 2 and

[58] The direct effect of Catholicism is also negative but about 20 times weaker than that of share of Protestants and not statistically significant.

3). Results also show, however, that the direct effect both of Protestant denomination and of education on work ethic is essentially negative. In fact, the positive effect of education for Protestants is not strong enough to offset education's direct negative effect on the value attached to work relative to spare time (0.010 vs. -0.033). Finally and most importantly, Schaltegger and Torgler's (2010) analysis reveals that the unemployed are actually statistically significantly more likely to put work first at the expense of spare time than are individuals with full-time jobs.

This latter result adds to the intriguing and counterintuitive negative associations between work ethic and practices and economic outcomes found by other work mentioned above (e.g. Bertrand and Mullainathan 2001; McCleary and Barro 2006; Weil 2009). A likely explanation for these findings is that work ethic data are prone to suffer the marginal preferences problem (Chapter 3), and this is borne out by a simple statistical analysis. We use the same WVS item as studied by Schaltegger and Torgler (2010), which reads as follows: "Do you agree or disagree with the following statements? Work should always come first, even if it means less spare time: 1 – Strongly agree; 2 – Agree; 3 – Neither agree or disagree; 4 – Disagree; 5 – Strongly disagree" (see European Values Study Group and World Values Survey Association 2006 and World Values Survey Association 2009). Figure 6.A.1 and Table 6.A.1 in the appendix depict our findings, showing that the unemployed typically report attaching more value to work than the full-time employed do; and that, similarly, societies in which unemployment rates are higher are also those societies in which work is valued higher. Overall, those who have little "work", i.e. the unemployed, appear to value work more highly than those with little leisure, i.e. the employed.

This problematic nature of the data used in most (if not all) previous quantitative analyses of the Weber thesis of course provides important inspiration to take up studying this thesis empirically using the EP method.

6.3. MEASURING THE WORK ETHIC: AN EXPERIENCED PREFERENCES APPROACH

6.3.1. Well-Being and Unemployment

There is a large literature in psychology showing and theoretically arguing that unemployment is associated with large psychic costs, even when its economic consequences

are controlled for (Darity and Goldsmith 1996; Haring et al. 1984; Hayes and Nutman 1981; McKee-Ryan et al. 2005). This literature can be traced back to over two centuries ago (Fryer 2001)—Eisenberg and Lazarsfeld (1938) is an early empirical study. In economics, Clark and Oswald (1994) analyze data from the British Household Panel and find that unemployed people have much lower levels of mental well being than those in work. Similar results have been reported by Clark (2003), Shields and Wheatley Price (2005), Stutzer and Lalive (2004), and Winkelmann and Winkelmann (1998), among others. What is more, it is clear that the effect runs from unemployment to SWB. Using longitudinal data, Clark et al. (2008) demonstrate that individuals who lose their job find their level of SWB decreasing substantially upon becoming unemployed, whilst they do not have lower SWB to begin with (see, also, Lucas et al. 2004). Finally, unemployment not only affects the well-being of the people losing their jobs, but also (i.e. simultaneously) has an indirect impact on the population as a whole (e.g. Di Tella et al. 2001, 2003; Di Tella and MacCulloch 2005).

6.3.2. The Protestant Work Ethic and the Psychic Cost of Unemployment

Our operationalization of work ethic in terms of the experienced preference for work builds on the SWB-(un)employment literature briefly surveyed above. Specifically, we use the degree to which having a job adds to people's experienced utility as an indicator of the importance people attach to work and, hence, of their work ethic. People attaching a lot of importance to work are hurt more by losing their job than people who think work is unimportant in life; this is what differences in the effect size of unemployment on experienced utility indicate. To test the Weber thesis we introduce religious heterogeneity in the adverse SWB effect of being unemployed (relative to having a job). Allowing the psychic cost of unemployment to vary across individuals from different denominations gives us the direct means to assess how much religious background matters for the value people attach to work.

Religion has been shown to be an important factor in mental health and SWB (Dezutter et al. 2006; Ellison 1995; Ellison and Fan 2008; Hackney and Sanders 2003; Koenig and Larson 2001; Pargament 2002). In addition there is evidence that the impact of economic factors on SWB differs between religious and non-religious groups. Clark and Lelkes (2005), for instance, find that religious beliefs may shield against part of the negative well-being effects of stressors like unemployment (see, also, Smith et al. 2003). Lelkes (2006) further reports that the effect of economic variables including income on happiness is smaller

among the religious than among the non-religious. Apparently, religious people value income less than non-believers.

To reach a conclusion concerning the Weber thesis, we look for differences between Protestants and non-Protestants. This results in the following hypothesis.

Hypothesis 1 *The subjective well-being of Protestants is influenced more by being unemployed (relative to having a job) than is the subjective well-being of people from other denominations.*

Hypothesis 1 is useful for testing one of the more common interpretations of the Weber thesis, namely that those being Protestants now are likely to have a stronger work ethic than people currently holding different religious beliefs. However, we have noted that Weber's original argument did not so much focus on Protestantism in the present but on a Protestant ethic as a historical factor, having evolved into a rational, secular "spirit of capitalism". A hypothesis closer to the original argument is therefore:

Hypothesis 2 *The subjective well-being of people from Protestant societies is influenced more by being unemployed (relative to having a job) than is the subjective well-being of people from other societies.*

We test these two hypotheses below. First, however, we describe the data and method used in our empirical analysis.

6.4. DATA AND METHOD

6.4.1. Description of the Data

The data we use in our empirical analysis is taken from the World Values Survey (WVS), all five waves (European Values Study Group and World Values Survey Association 2006; World Values Survey Association 2009). Since we have used these data before in Chapter 3 and Chapter 5, we do not describe their source here. Below we do discuss the variables that we use, however. Where needed, we supplement the WVS data with aggregate-level data from other sources. Details follow.

Dependent Variable

The dependent variable of interest concerns individuals' life satisfaction, measured by the item asking "All things considered, how satisfied are you with your life as a whole these days?", with the answer scale ranging from "1-Dissatisfied" to "10-Satisfied." We have previously used this SWB indicator in Chapter 5 and we also discussed it in Chapter 4. To facilitate the ease of interpretation of the findings and following a large literature in psychology, we analyze this SWB variable as though it is an ordinal variable (range 1 to 10), noting that this will not substantially affect our results (see, for example, Clark et al. 2008, Ferrer-i-Carbonell and Frijters 2004, Frey and Stutzer 2000, and Stevenson and Wolfers 2008).[59] In our analyses we include only individuals with non-missing satisfaction scores.

Key Independent Variables

The independent variables we are most interested in are individuals' employment status and religious denomination. The WVS item on employment status discerns eight categories: "Full-time employed," "Part time" employed, "Self employed," "Retired," "Housewife,", "Student," "Unemployed," and "Other". We construct dummies for seven of these eight categories, taking full-time employed as the base category. For religious denomination we use the answer to the question whether people belong to a religious denomination and, if yes, to which one they belong. The WVS codes a great number of possible denominations, including being Protestant. In our analyses we include only individuals who indicated they belong to a religious denomination and create a dummy with value 1 if the respondent is Protestant and 0 if he or she belongs to another denomination. To test Hypothesis 1—Protestants value work higher than individuals from other denominations do—we interact the Protestant dummy with the unemployment dummy. The empirical analysis then shows how much more (or less) Protestants suffer from being unemployed (relative to full-time employed) than other denominations do.

For testing Hypothesis 2—unemployment hurts more in Protestant societies—we supplement the WVS data with data on Protestant societies from Norris and Inglehart (2004: 14). Applying the Norris-Inglehart classification we identify the following countries in our sample as Protestant: Australia, Estonia, Finland, Germany, Latvia, the Netherlands, New

[59] Note that the same cardinality assumption was implicit in our empirical assessment of the consistency of EP measures of preference heterogeneity across different types of well-being assessments in Chapter 5. As a robustness check, below we also estimate our main empirical models using the measure of happiness included in the WVS. Again see Chapters 4 and 5 for details about this SWB indicator.

Zealand, Norway, South Africa, Zimbabwe, Sweden, Switzerland, Uganda, United Kingdom, and the United States. Applied to individuals, respondents get a dummy score 1 if they live in one of these countries and score 0 otherwise. For our robustness checks we also experiment with a dummy indicating whether a country is Protestant based on a contemporary numerical majority criterion. Using the 50%-benchmark, we identify Finland, Great Britain, Norway, South Africa, Sweden, and the United States as societies in our sample in which Protestantism currently is the dominant religion (classification based on data from the CIA Factbook; Central Intelligence Agency 2009). As with Hypothesis 1, the empirical model used to test Hypothesis 2 includes an interaction term, which captures how much more (or less) individual unemployment lowers SWB in Protestant societies than in non-Protestant societies.

Other Independent Variables

Our empirical models include several other independent variables, both at the individual and at the societal level. The WVS includes different items probing respondents about their background such as their health status and income scale. Controlling for these situational factors is important as unemployment is likely to have an indirect impact on SWB through its effect on these individual circumstances. Notably, unemployment is associated with lower income. The negative happiness effect of unemployment may also be driven by other situational factors associated with unemployment. People in poor health, for example, likely have a higher risk of being unemployed so that the effect of unemployment on happiness may partly be a result of employment status proxying for health status.

Accordingly, we include the following variables as control variables in our analysis: sex (1 = male), age and age squared, income scale (1-10), education (Inadequately completed elementary education, Completed (compulsory) elementary education, Incomplete secondary school: technical/vocational type, Complete secondary school: technical/vocational type, Incomplete secondary: university-preparatory type, Complete secondary: university-preparatory type, Some university without degree, or University with degree), state of health (1, very poor – 5, very good), marital status (Married, Living together as married, Divorced, Separated, Widowed, or Never married). With the exception of health status, these are the same individual-level variables that we used in the empirical analysis of the marginal preferences problem and the (alleged) democracy paradox of Islam in Chapter 3. All these variables have further been used in our assessment of the psychometric quality of EP measures in the previous chapter. We treat these independents the same as our dependent variable so that individuals with missing answers, or "unanswered" or "don't know" response

on the relevant variable are excluded. This way, for our basic model, we end up with data on 149,462 individuals from 82 societies (several of the latter sampled more than once).[60] Table 6.1 gives a short description and some summary statistics for the main variables included in our analysis. When we check the robustness of our results we also look at possible moderating effects of some additional variables, specifically whether religiosity and being raised religiously affect the psychic cost of unemployment. Since data on these variables are missing for many individuals we do not present descriptives for them in Table 6.1, but see Table 6.A.2 in the appendix.

Table 6.1: Descriptive Statistics for Main Variables.

Variable and description	Mean and standard deviation
Dependent variable	
Satisfaction (1-10)	6.43 (2.54)
Independent variables	
Unemployed (1 = yes)	9.9% (29.9%)
Protestant denomination (1 = yes)	13.5% (34.2%)
Unemployed & Protestant (interaction term)	1.4% (11.7%)
Protestant society (1 = yes)	17.9% (38.3%)
Unemployed in a Protestant society (interaction term)	1.8% (13.2%)
Employment status	
Full-time employed [base category]	32.5% (46.9%)
Part-time employed	7.4% (26.1%)
Self-employed	12.0% (32.5%)
Retired	11.6% (32.0%)
Housewife	16.8% (37.4%)
Student	7.6% (26.5%)
Other	2.1% (14.5%)
Sex (1 = male)	47.7% (49.9%)
Age	40.4 (15.8)
Income scale (1-10)	4.48 (2.36)
State of health (1, very poor – 5, very good)	3.81 (.887)

[60] The 82 societies covered, are: Albania, Andorra, Azerbaijan, Argentina, Australia, Bangladesh, Armenia, Bosnia and Herzegovina, Brazil, Bulgaria, Belarus, Canada, Chile, China, Taiwan, Colombia, Cyprus, Czech Republic, Dominican Republic, El Salvador, Ethiopia, Estonia, Finland, France, Georgia, Germany, Ghana, Guatemala, Hong Kong, India, Indonesia, Iran, Iraq, Italy, Japan, Jordan, South Korea, Kyrgyzstan, Latvia, Lithuania, Malaysia, Mali, Mexico, Moldova, Morocco, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Peru, Philippines, Poland, Puerto Rico, Romania, Russian Federation, Rwanda, Saudi Arabia, Slovakia, Viet Nam, Slovenia, South Africa, Zimbabwe, Spain, Sweden, Switzerland, Thailand, Trinidad and Tobago, Turkey, Uganda, Ukraine, Macedonia, Egypt, Great Britain, Tanzania, United States, Burkina Faso, Uruguay, Venezuela, Serbia and Montenegro, Zambia, and Serbia (see Note 53).

Table 6.1, continued.

Variable and description	Mean and standard deviation
Marital status	
Married [base category]	59.2% (49.1%)
Living together as married	6.2% (24.1%)
Divorced	2.9% (16.7%)
Separated	1.7% (12.8%)
Widowed	6.2% (24.0%)
Never married	23.9% (42.7%)
Education	
Inadequately completed elementary education	14.7% (35.4%)
Completed (compulsory) elementary education	15.1% (35.8%)
Incomplete secondary: technical/vocational	7.4% (26.1%)
Complete secondary: technical/vocational	16.8% (37.4%)
Incomplete secondary: university-preparatory	8.6% (28.1%)
Complete secondary: university-preparatory	16.4% (37.0%)
Some university without degree	6.5% (24.7%)
University with degree [base category]	14.5% (35.2%)
Society with present-day Protestant domination	11.2% (31.5%)
Aggregate unemployment	10.6% (7.4%)

Our robustness checks further include some country-level variables thought to affect how much it hurts to be unemployed relative to having a full-time job. Notably we find that social security benefits likely affect the disutility of unemployment. To control for a possible bias as a result of this, we extend our baseline models to include the government share in total consumption expenditure (which proxies for the extent of the welfare state) as a moderator. Data come from the World Development Indicators (World Bank 2010) and Table 6.A.2 gives descriptive statistics for this variable too.[61]

Finally, next to these robustness checks, we also extend our basic models to explore two possible channels through which being Protestant or living in a Protestant society may affect the psychic cost of unemployment. The first of these concerns materialism, where the idea is that materialism causes a greater preference for work not for intrinsic reasons but for the extrinsic motivation of the financial rewards associated with having a job. Secondly, it is possible that informal institutions, particularly, peer pressure due to stigma effects of unemployment (Clark 2003; Clark and Oswald 1994; Shields and Wheatley Price 2005; Stutzer and Lalive 2004) affect the psychic distress associated with unemployment. We assess this possibility by examining whether the general unemployment rate moderates the SWB

[61] As an alternative, we also estimate models that allow for moderating effects of the type of welfare state, as classified by Hall and Soskice (2001) (H&S) and Esping-Andersen (1989, 1990) (A-E). Again see Table 6.A.2 for descriptives.

effect of unemployment. To ensure that we have as many observations as possible, we measure aggregate unemployment using the WVS employment status item discussed above (see Table 6.1).

6.4.2. Model and Estimation

As discussed in Chapter 3, an important feature of the WVS data is that they are structured hierarchically with individual observations (Level 1) nested in countries (Level 2). This means that we again apply multilevel modeling (Gelman and Hill 2007; Raudenbush and Bryk 2002; Snijders and Bosker 1999). Next to taking account of the clustering of observations (Moulton 1986, 1990; Wooldridge 2003) and other advantages (ibidem; see Chapter 3), an important reason to use multilevel modeling for our current analysis is that it is well-suited for dealing with cross-level interactions. Similar to the model in Chapter 3, our current model has explanatory variables at two levels. It also includes a term that allows the psychic cost of unemployment to vary between Protestant and non-Protestant societies, however, and traditional statistical techniques have difficulty in properly handling this cross-level interactions.

Some other studies applying multilevel modeling to the analysis of individual SWB scores across different geographical regions are by Pittau et al. (2010) and by Subramanian et al. (2005). These studies investigate the determinants of happiness of persons residing in different regions and use multilevel modeling to separate the impact of individual factors such as income and education from the independent impact of individuals' community. More closely related to our research is the work by Hui et al. (2004). They use WVS data to analyze whether societal scores on Hofstede's (1980) power distance index, which reflect the extent to which people in a particular society expect and accept that power is distributed unequally, moderate the effect of empowerment on job satisfaction. Their analysis shows that higher power distance lowers the positive effect of empowerment on job satisfaction and may even cancel out the positive association altogether.

For the formal empirical model we have an individual i (Level 1) who resides in country j (Level 2). The individual's subjective well-being is denoted by SWB_{ij}. The independent variables of interest are the unemployment dummy U_{ij} (1 if yes), a Protestant dummy P_{ij} (1 if yes), a dummy S_j that equals 1 if the society is predominantly Protestant and 0 otherwise. Testing our hypotheses requires inclusion of two interaction terms: $U_{ij} \times P_{ij}$ and

$S_j \times U_{ij}$. We also include x_{ij}, the set of other individual determinants of SWB, and z_j, the set of country-level variables. This yields the following general, level-1, within-country model:

$$SWB_{ij} = \beta_{0j} + \beta_{1j}U_{ij} + \beta_{2j}P_{ij} + \beta_{3j}\left(U_{ij} \times P_{ij}\right) + \beta_{4j}x_{ij} + \varepsilon_{ij}.$$

Similarly, the level-2, between-country model is given by:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}S_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}S_j + u_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}z_j + u_{2j}\,. \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}z_j + u_{3j} \\
\beta_{4j} &= \gamma_{40} + \gamma_{41}z_j + u_{4j}
\end{aligned}$$

Combining the two gives our complete model:

$$\begin{aligned}
SWB_{ij} = {} & \gamma_{00} + \gamma_{01}S_j + \gamma_{10}U_{ij} + \gamma_{20}P_{ij} + \gamma_{30}\left(U_{ij} \times P_{ij}\right) + \gamma_{40}x_{ij} \\
& + \gamma_{11}\left(S_j \times U_{ij}\right) + \gamma_{21}z_jP_{ij} + \gamma_{31}z_jU_{ij} \times P_{ij} + \gamma_{41}z_jx_{ij} \quad . \qquad (6.1) \\
& + \left[u_{0j} + u_{1j}U_{ij} + u_{2j}P_{ij} + u_{3j}U_{ij}P_{ij} + u_{4j}x_{ij} + \varepsilon_{ij}\right]
\end{aligned}$$

In this model the moderating effect of Protestantism and living in a historically Protestant society on the SWB effect of unemployment captures the distinct Protestant work ethic (if such a thing exists at all). To test our hypotheses, we look at the coefficients γ_{30} and γ_{11} as these show both whether there is a specific Protestant component to the value of having a job and what the size of this effect is. We have further specified the model broadly to incorporate other cross-level interactions, mainly with an eye to checking the robustness of our basic findings. The terms u_{1j} to u_{4j} thereby capture country-specific deviations from the mean SWB effect of being unemployed, having Protestant denomination, et cetera. These random slopes are of no concern to us, however, as we are interested in group heterogeneity driven by religious background only. Because we also include a constant (γ_{00}), the general model is a varying-intercepts, varying-coefficients (or random-intercepts, random-slopes) model, where u_{0j} denotes country-specific deviations from the average intercept in our sample. The term in square brackets constitutes the random part of the complete model and the other terms the fixed part. We estimate the model using maximum likelihood.

6.5. EMPIRICAL RESULTS

6.5.1. Basic Results

We first test Hypothesis 1, investigating the impact of being Protestant on individuals' evaluation of unemployment. Does the structure of SWB differ between religious denominations? Our results indicate that this is indeed the case (Table 6.2). Models 1 and 2 show that, in accordance with the literature, unemployment has a robust, negative effect on well-being. Being of Protestant denomination, as opposed to another denomination, is associated with higher SWB, adding about 0.07 to satisfaction on average (Model 3). The effect we are primarily interested in, however, is that of being both Protestant and unemployed. Upon including this interaction term, we find it has a negative coefficient of almost 0.13, indicating that whereas unemployment reduces experienced utility regardless of religious denomination, it has an additional negative effect for Protestants (Model 4). In percentage terms, individual Protestantism adds about 40% to the psychic cost of unemployment (0.129/0.325). This effect is statistically highly significant. Hence, it appears strongly that at the individual level, unemployment hurts Protestants much more than it does non-Protestants. This is a preliminary confirmation of the Weber thesis.

Table 6.2: The Psychic Cost of Unemployment for Protestants.

	Model 1	Model 2	Model 3	Model 4
Intercept	6.53***	4.07***	4.06***	4.06***
	(.124)	(.128)	(.128)	(.128)
Unemployed	-.652***	-.343***	-.342***	-.325***
	(.022)	(.022)	(.022)	(.023)
Protestant denomination	-	-	.053**	.067***
			(.021)	(.021)
Interaction				
Unemployed & Protestant	-	-	-	-.129**
				(.055)
Employment status				
Part-time employed	-.200***	-.082***	-.082***	-.082***
	(.024)	(.023)	(.023)	(.023)
Self-employed	-.151***	.003	.003	.003
	(.021)	(.020)	(.020)	(.020)
Retired	-.327***	.068***	.068***	.068***
	(.021)	(.025)	(.025)	(.025)
Housewife	-.047**	.182***	.182***	.183***
	(.018)	(.020)	(.020)	(.020)

Table 6.2, continued.

	Model 1	Model 2	Model 3	Model 4
Student	.157***	.063**	.063**	.062**
	(.024)	(.026)	(.026)	(.026)
Other	-.343***	-.048	-.048	-.048
	(.044)	(.042)	(.042)	(.042)
Control variables				
Income scale	-	.170***	.170***	.170***
		(.003)	(.003)	(.003)
Health	-	.655***	.655***	.655***
		(.007)	(.007)	(.007)
Living together as married	-	-.098***	-.098***	-.097***
		(.026)	(.026)	(.026)
Divorced	-	-.384***	-.384***	-.384***
		(.035)	(.035)	(.035)
Separated	-	-.555***	-.554***	-.554***
		(.045)	(.045)	(.045)
Widowed	-	-.259***	-.259***	-.259***
		(.026)	(.026)	(.026)
Never married	-	-.238***	-.238***	-.238***
		(.018)	(.018)	(.018)
Further controls	No	Yes	Yes	Yes
-2Loglikelihood	671,266.4	655,082.0	655,075.4	655,069.9

Notes: Standard errors in parentheses. * (**) (***) indicates significance at the 10% (5%) (1%) level. Sample consists of 149,462 individuals residing in 82 societies. All models include varying intercepts. Unreported control variables are education, sex, age and age squared. For "Protestant denomination" the reference category is formed by all other religious denominations. The base category is a married, non-Protestant woman with a university degree who is full-time employed.

What about the societal level? As noted, Weber's original thesis refers to a capitalist spirit that has its roots in a Protestant ethic, but which, over time, has grown into an independent, even secular worldview. On the basis of Weber's work we would therefore not so much expect a link between one's work ethic and being a Protestant now, as a link between one's work ethic and whether one lives in a society historically dominated by the Protestant ethic and having developed a "capitalist spirit." Hypothesis 2, stating that there should be a relation between the effect size of unemployment on SWB and living in a society in which Protestantism is the dominant religion, is therefore a more appropriate test of the original Weber thesis.

Table 6.3 presents the results for our test of Hypothesis 2. As Model 6 shows, being unemployed hurts people from Protestant dominated societies significantly more than others (-0.329). Moreover, when this cross-level interaction term is included, we find that individual Protestantism does not significantly contribute to the psychic cost of unemployment anymore,

except within Protestant societies where it lowers SWB by an additional 0.248 points (Model 7). Model 8 shows this result holds independent of the inclusion of the second-order interaction terms (being Protestant in a Protestant society, and being unemployed and Protestant in a Protestant society).[62] The Weber thesis thus seems confirmed, while apparently it is living in a Protestant society rather than being Protestant oneself that matters most.

Table 6.3: Protestantism and the Psychic Cost of Unemployment.

	Model 5	Model 6	Model 7	Model 8
Unemployed	-.342*** (.022)	-.287*** (.023)	-.297*** (.024)	-.288*** (.024)
Level 1				
Protestant	.052** (.021)	.050** (.021)	.018 (.029)	.049** (.022)
Interaction				
Unemployed & Protestant	-	-	.130* (.079)	.010 (.060)
Level 2				
Protestant society	.376 (.289)	.405 (.289)	.384 (.289)	.406 (.289)
Cross-level interaction				
Unemployed in a Protestant society	-	-.329*** (.052)	-.248*** (.066)	-.333*** (.056)
Other interactions				
Protestant in a Protestant society	-	-	.073* (.043)	-
Unemployed & Protestant in a Protestant society	-	-	-.288** (.121)	-
-2Loglikelihood	655,073.8	655,033.3	655,026.6	655,033.3

Notes: See Table 6.2. Full set of controls included. The base category is a married, non-Protestant woman with a university degree who is full-time employed and lives in a non-Protestant country.

6.5.2. Robustness

A chief potential objection to our conclusions is that, for some reason, the life satisfaction question we use is not a good measure of SWB, invalidating our results. In order to test for this possibility, we replace the life satisfaction measure with an alternative indicator of SWB to see if our results stand up. Using individuals' happiness ratings as the independent variable

[62] To be sure, the model without the other interaction terms included (Model 8) is not our preferred model. Nevertheless, we present it to address beforehand possible questions about the role of these second-order terms in our main result that the dominant effect of Protestantism on the psychic cost of unemployment is at the societal level and not at the individual level.

in our empirical model does not change our findings (see Table A.6.3 in the appendix). Confirming our earlier results, we find the negative effect of unemployment on self-reported happiness to be twice as strong for Protestants than it is for the non-Protestant population (the additional effect being -0.065 with a standard error of 0.017). Similarly, at the macro-level, we observe that the impact of unemployment on happiness is significantly stronger in Protestant societies (additional effect -0.100, standard error 0.017). Just as in the original analysis, the effect of individual Protestantism disappears when including Protestantism at the societal level, again suggesting that it is living in a Protestant society rather than being Protestant oneself that matters for the preference for work.

Our second robustness check assesses the sensitivity of our results to the classification of societies as Protestant. Replacing the old classification based on Norris and Inglehart (2004) with the classification based on present-day dominance does not alter our results (see Table A.6.4 in the appendix). Unemployment still hurts significantly more in Protestant societies, the additional effect being -0.257 with a standard error of 0.073.

In addition to robustness to changes in measures, we have also checked whether our results are altered when including additional variables. The most prominent of these is the extent of social security in a society. Since generous welfare states tend to mitigate the negative (income) effects of unemployment, the above results may be biased. In particular, if predominantly Protestant countries are disproportionately Anglo-Saxon and characterized by less elaborate welfare states, the higher experienced disutility of unemployment in Protestant societies may simply reflect objectively higher cost of unemployment in Protestant societies rather than any differences in work ethic. In order to test this possibility, we add government expenditure as a share of GDP to the analysis as a proxy for the extent of the welfare state (Table 6.4).

Our results are robust for controlling for differences in the size of the welfare state. The effect of unemployment on SWB does not seem to differ between countries with a more or less extensive government role. More importantly, the effect of Protestantism on the psychic cost of unemployment remains unaffected. Protestants and those living in a Protestant society continue to experience higher losses in experienced utility from unemployment than others do (Models 11-12). This result remains when we switch to other proxies for the size of the welfare state, based on Hall and Soskice (2001) or Esping-Anderson (1989, 1990) (see Tables 6.A.5a and 6.A.5b in the appendix). We conclude that our finding that being unemployed hurts much more in countries in which Protestantism is the dominant religion is not driven by differences in welfare states.

Table 6.4: Government Consumption and the Psychic Cost of Unemployment.

	Model 9	Model 10	Model 11	Model 12
Unemployed	-.366*** (.023)	-.363*** (.023)	-.344*** (.024)	-.306*** (.025)
Level 1				
Protestant	.060*** (.022)	.060*** (.022)	.057** (.024)	.035 (.022)
Interactions				
Unemployed & Protestant	-	-	-.105 (.066)	-
Level 2				
Protestant society	.809** (.310)	.808** (.310)	.793** (.310)	-.783** (.356)
Government consumption share	-7.85*** (.635)	-7.77*** (.636)	-7.90*** (.638)	-10.1*** (.660)
Cross-level interaction				
Unemployed in a Protestant society	-	-	-	-.521*** (.110)
Government consumption share × unemployed	-	-.742* (.417)	-.397 (.441)	.119 (.468)
Other interactions				
Government consumption share × Protestant	-	-	1.28** (.525)	-
Government consumption share × (Unemployed & Protestant)	-	-	-1.92 (1.44)	-
Government consumption share × Protestant society	-	-	-	34.2*** (2.69)
Government consumption share × (Unemployed in a Protestant society)	-	-	-	4.60** (2.15)
-2Loglikelihood	578,893.5	578,890.4	578,878.2	578,682.3

Notes: See Table 6.3. Government consumption share is mean-centered.

Finally, we might speculate that any effect of Protestantism on work ethic would be moderated by the degree of religiosity of people. In order to check this, we include a measure of religiosity and investigate how it interacts with Protestantism and unemployment (Table 6.5). As expected from a reading of the existing literature (Dezutter et al. 2006; Ellison 1995; Ellison and Fan 2008; Hackney and Sanders 2003; Koenig and Larson 2001; Pargament 2002), we find that religiosity has a direct, positive effect on SWB (Model 13). But this does not affect the impact of Protestantism on the experienced costs of unemployment. More interestingly, all but one of the interactions with religiosity are statistically insignificant. The one exception is the interaction "unemployed & not religious in a Protestant society," which has a positive effect. This latter result is in line with our earlier finding that within Protestant societies individual Protestantism increases the adverse SWB effect of unemployment. All in

all, however, these results indicate that differences in religiosity do not add much to our model.

Table 6.5: The Role of Personal Religiosity.

	Model 13	Model 14	Model 15	Model 16
Unemployed***	-.344 (.022)	-.343 (.024)	-.321 (.025)	-.282 (.026)
Level 1				
Protestant	.058*** (.022)	.053** (.021)	.074*** (.023)	.051** (.021)
Religious person				
Not a religious person***	-.222 (.018)	-.225 (.017)	-.219 (.018)	-.222 (.019)
A convinced atheist***	-.389 (.061)	-.393 (.056)	-.390 (.060)	-.384 (.064)
Interactions				
Unemployed & Protestant	-	-	-.158*** (.060)	-
Unemployed & not a religious person	-	.007 (.051)	-.020 (.055)	-.030 (.057)
Unemployed & a convinced atheist	-	-.155 (.171)	-.170 (.183)	-.029 (.194)
Unemployed & not a religious person & Protestant	-	-	.189 (.159)	-
Unemployed & a convinced atheist & Protestant	-	-	.103 (.513)	-
Protestant & not a religious person	-.036 (.053)	-	-.047 (.048)	-
Protestant & a convinced atheist	.068 (.179)	-	-.017 (.164)	-
Level 2				
Protestant society	.400 (.291)	.398 (.291)	.398 (.291)	.432 (.291)
Cross-level interactions				
Unemployed in a Protestant society	-	-	-	-.366*** (.058)
Unemployed & not religious in a Protestant society	-	-	-	.228* (.133)
Unemployed & a convinced Atheist in a Protestant society	-	-	-	-.479 (.411)
Not religious in a Protestant society	.011 (.047)	-	-	-.018 (.042)
A convinced Atheist in a Protestant society	-.119 (.142)	-	-	-.039 (.130)
-2Loglikelihood	634,566.4	634,566.8	634,559.5	634,521.3

See Table 6.3.

6.5.3. Extensions: Income and Social Stigma Effects as Possible Channels

The previous analyses show that Protestantism increases the negative effect of unemployment on experienced utility, a finding that is robust for changes in measures and controlling for several factors. The question that remains is whether this result is indicative of a direct effect of a Protestant ethic on the costs of unemployment or that it runs through other, indirect channels. One such potentially important channel is income. An explanation for the finding that unemployment seems to hurt Protestants more could be that they are more materialistic and care more about the income associated with having a job. In order to investigate this possibility, we have included the effect of income for Protestants and the remainder of the population in our estimation of heterogeneous experienced utility functions (Table 6.6).

Table 6.6: Protestantism versus Materialism.

	Model 17	Model 18	Model 19	Model 20
Unemployed***	-.343 (.022)	-.323 (.023)	-.344 (.022)	-.284 (.023)
Level 1				
Protestant	.128*** (.037)	.155*** (.039)	.053*** (.021)	.051** (.021)
Income scale***	.173 (.003)	.173 (.003)	.174 (.003)	.175 (.003)
Interactions				
Unemployed & Protestant	-	-.149*** (.055)	-	-
Income scale × Protestant	-.016** (.007)	-.018*** (.007)	-	-
Level 2				
Protestant society	.377 (.289)	.375 (.289)	.465 (.291)	.528* (.291)
Cross-level interaction				
Unemployed in a Protestant society	-	-	-	-.362*** (.052)
Income scale × Protestant society	-	-	-.019*** (.006)	-.026*** (.006)
-2Loglikelihood	655,067.9	655,060.7	655,064.2	655,016.3

Notes: See Table 6.3.

We find that it is not because they care more about income that Protestants are hurt more by joblessness. In fact, whereas higher incomes contribute significantly to satisfaction, they do much less so for Protestants. This shows that if anything, Protestants care less about

income than non-Protestants. The higher valuation of employment by Protestants is apparently rooted in a stronger intrinsic appreciation of work.

Table 6.7: Protestantism, Social Stigma, and the Psychic Costs of Unemployment.

	Model 21	Model 22	Model 23	Model 24
Unemployed	-.331*** (.022)	-.340*** (.024)	-.327*** (.024)	-.295*** (.024)
Level 1				
Protestant	.043** (.021)	.044** (.021)	.061*** (.022)	.044** (.021)
Interactions				
Unemployed & Protestant	-	-	-.134* (.073)	-
Level 2				
Protestant society	.352 (.273)	.351 (.273)	.347 (.274)	.431 (.275)
Aggregate unemployment	-2.10*** (.260)	-2.13*** (.261)	-2.11*** (.264)	-3.56*** (.328)
Cross-level interaction				
Unemployed in a Protestant society	-	-	-	-.397*** (.076)
Aggregate unemployment × unemployed	-	.224 (.226)	.353 (.262)	.853*** (.286)
Other interactions				
Aggregate unemployment × Protestant	-	-	-.151 (.214)	-
Aggregate unemployment × (Unemployed & Protestant)	-	-	-.005 (.558)	-
Aggregate unemployment × Protestant society	-	-	-	3.95*** (.547)
Aggregate unemployment × (Unemployed in a Protestant society)	-	-	-	-.234 (.558)
-2Loglikelihood	655,008.8	655,007.9	655,000.8	654,907.1

Notes: See Table 6.3. Aggregate unemployment is mean-centered.

A second channel is the social stigma involved in unemployment in some societies. In societies in which the stigma attached to unemployment is relatively large, the well-being effect of unemployment is likely to be larger. If for some reason in Protestant countries unemployment carries more of a social stigma, this may very well drive our results. One such reason may be that unemployment rates are, on average, lower in Protestant societies than elsewhere. According to the literature, a high unemployment rate implies that being unemployed is a relatively "normal" condition which one shares with a large part of society. In low-unemployment societies, by contrast, being unemployed is relatively rare, which adds

to the unemployed's marginalization. To control for this, we have added a country's unemployment rate to our analysis (Table 6.7). Doing so, we find that aggregate unemployment negatively affects SWB. What is more, in line with our expectations and previous studies (Clark 2003; Clark and Oswald 1994; Shields and Wheatley Price 2005; Stutzer and Lalive 2004), we find that a high rate of aggregate unemployment mitigates the psychic cost of unemployment experienced by individuals (Model 24). But none of this affects our main finding that Protestants and people living in Protestant societies seem to care more about unemployment than others. Protestantism continues to have a strong effect on the preference for work.

6.6. DISCUSSION AND CONCLUSION

Ever since Max Weber's seminal work on Protestantism and the spirit of capitalism started a debate about the impact of religion on economic performance, religion has featured prominently in comparative research on values. This chapter examines whether Protestantism spawned a distinctive work ethic. We apply the EP method and operationalize work values by examining the impact of religion on the involuntariness of unemployment. To test Weber's Protestant work ethic thesis we compare the losses in experienced utility due to being unemployed between individuals of Protestant and non-Protestant denomination. We argue that, if this thesis is correct and Protestants / people from Protestant societies indeed have a relatively stronger work ethic, joblessness should hurt these groups more than it does others. For our empirical analysis we use data from the WVS, covering 82 societies and comprising almost 150,000 individuals. Results show that unemployment has a negative effect on SWB in general, but that Protestantism makes this effect much larger, establishing the contemporary relevance of Weber's insights on a specific Protestant work ethic more than a century after their inception. This result is robust for inclusion of welfare state regimes, income effects, unemployment levels and a range of individual-level control variables.

Interestingly, the effect appears stronger at the country level: it is not so much Protestant individuals who are hurt more by being unemployed as it is individuals (both Protestants and non-Protestants) living in Protestant societies. Individual Protestant denomination aggravates the psychic cost of unemployment for individuals (i.e. Protestants) living in a Protestant society. On the basis of this societal effect of Protestantism on work ethic we find that the channel linking religion to differences in values, specifically work ethic,

is very likely a subtle one. Studies exploring religion's economic consequences might benefit from incorporating this nuance and investigating it further.

APPENDIX TO CHAPTER 6

Table 6.A.1: Work versus Spare Time by Employment Status.

Employment status	Distribution of responses [%]					Mean
	Strongly agree (1)	Agree (2)	Neither agree or disagree (3)	Disagree (4)	Strongly disagree (5)	
Full-time employed [48,696]	21.1	33.8	16.9	22.4	5.8	2.58 (1.21)
Unemployed [13,607]	29.5	36.8	15.3	14.8	3.6	2.26 (1.14)

Notes: Standard deviations in parentheses and number of observations in square brackets. The sample covers 80 countries. Due to rounding, category percentages may not add up to 100%. A chi-square test rejects the null-hypothesis that the distribution of the responses by the full-time employed and the unemployed are the same ($p = 0.000$). A t-test shows that the cardinal difference in valuation (0.32) is statistically highly significant ($p = 0.000$).

Table 6.A.2: Descriptive Statistics for Selected Variables.

Variable and description [Number of cases / societies]	Mean and standard deviation
Dependent variable	
Happiness [150,867 / 82]	3.04 (.764)
Independent variables	
Religious person [144,823 / 82]	
A religious person [base category]	81.8% (38.6%)
Not a religious person	17.0% (37.5%)
A convinced atheist	1.2% (11.0%)
Raised religiously (1 = yes) [49,245 / 42]	74.0% (43.9%)
Government share in consumption [132,149 / 74]	14.9% (5.1%)
Type of welfare state I (H&S)	
Not classified [base category]	87.3% (33.3%)
Liberal Market Economy	5.7% (23.3%)
Coordinated Market Economy	7.0% (25.5%)
Type of welfare state II (E-A)	
Not classified [base category]	86.6% (34.0%)
Liberal	5.7% (23.3%)
Conservative	5.3% (22.4%)
Social Democratic	2.4% (15.2%)

Notes: See Table 6.1. Religious person has a fourth category, "Other", which we have removed because it contains only 6 observations (0.004% of the original total of 144,829). The 42 societies with data on religious upbringing are: Albania, Azerbaijan, Argentina, Australia, Bangladesh, Armenia, Bosnia and Herzegovina, Brazil, Bulgaria, Belarus, Chile, Taiwan, Colombia, Czech Republic, Dominican Republic, Estonia, Finland, Germany, India, Latvia, Lithuania, Mexico, Moldova, New Zealand, Nigeria, Norway, Peru, Puerto Rico, Romania, Russian Federation, Slovakia, South Africa, Spain, Sweden, Switzerland, Turkey, Ukraine, Macedonia, United States, Uruguay, Venezuela, and Serbia and Montenegro. Hall and Soskice (H&S) (2001) classify the following countries as liberal market economies: Australia, Canada, Ireland, New Zealand, United Kingdom, and the United States; and Austria, Belgium, Denmark, Finland, Germany, Iceland, Japan, Netherlands, Norway, Sweden, Switzerland as coordinated market economies. According to the classification by Esping-Andersen (E-A) (1989, 1990), Australia, Canada, Ireland, New Zealand, United Kingdom, and the United States are liberal welfare states, France, Italy, Finland, Germany, Japan, and Switzerland are conservative welfare states, and Austria, Belgium, Denmark, the Netherlands, Norway, and Sweden are social-democratic welfare states.

Table 6.A.3: Protestantism, Unemployment, and Self-Reported Happiness.

	Model A1	Model A2	Model A3	Model A4
Intercept	2.13*** (.035)	2.13*** (.035)	2.13*** (.035)	2.13*** (.035)
Unemployed	-.068*** (.007)	-.059*** (.007)	-.051*** (.007)	-.049*** (.007)
Level 1				
Protestant	.032*** (.006)	.039*** (.007)	.032*** (.006)	.032*** (.009)
Interaction				
Unemployed & Protestant	-	-.065*** (.017)	-	-.027 (.024)
Level 2				
Protestant society	.072 (.069)	.071 (.068)	.081 (.068)	.078 (.069)
Cross-level interaction				
Protestant society & unemployed	-	-	-.100*** (.016)	-.088*** (.020)
Other interactions				
Protestant in a Protestant society	-	-	-	.006 (.013)
Unemployed & Protestant in a Protestant society	-	-	-	-.003 (.037)
-2Loglikelihood	304,795.8	304,780.7	304,756.4	304,753.8

Notes: See Table 6.3. Data concern 150,867 individuals from 82 societies (see Table A.6.2).

Table 6.A.4: Robustness of the Psychic Cost of Unemployment in Protestant Societies.

	Model A5	Model A6	Model A7
Unemployed	-.342***	-.299***	-.302***
	(.022)	(.023)	(.024)
Level 1			
Protestant	.052**	.051**	.063**
	(.021)	(.021)	(.026)
Interaction			
Unemployed & Protestant	-	-	.037
			(.073)
Level 2			
Society with present-day Protestant domination	.889**	.915**	.925**
	(.410)	(.410)	(.410)
Cross-level interaction			
Unemployed in a society with present-day Protestant domination	-	-.308***	-.257***
		(.056)	(.073)
Other interactions			
Protestant in a society with present-day Protestant domination	-	-	-.030
			(.047)
Unemployed & Protestant in a society with present-day Protestant domination	-	-	-.144
			(.125)
-2Loglikelihood	655,070.8	655,040.3	655,037.7

Notes: See Table 6.3.

Table 6.A.5a: Well-Being, Unemployment, and Protestantism in the Welfare State (H&S).

	Model A8	Model A9	Model A10	Model A11
Unemployed	-.342*** (.022)	-.341*** (.022)	-.324*** (.023)	-.293*** (.024)
Level 1				
Protestant	.052** (.021)	.054*** (.021)	.057** (.026)	.051 (.021)
Interactions				
Unemployed & Protestant	-	-	-.149** (.060)	-
Level 2				
Protestant society	-.499 (.350)	-.498 (.350)	-.498 (.350)	-.770* (.415)
Liberal market economy	1.12** (.458)	1.19** (.458)	1.17** (.458)	.956 (.640)
Coordinated market economy	1.58*** (.452)	1.59*** (.452)	1.58*** (.452)	.528 (.900)
Cross-level interactions				
Unemployed in a Protestant society	-	-	-	-.349*** (.058)
Unemployed in a Liberal market economy	-	.244** (.109)	.234* (.128)	.220* (.125)
Unemployed in a Coordinated market economy	-	-.275** (.109)	-.365*** (.135)	-.670 (.657)
Other interactions	-	-	Included	Included
-2Loglikelihood	655,060.1	655,048.2	655,039.9	655,009.2

Notes: See Table 6.3. So as not to clutter the table we do not report results for the other interactions terms (Protestant in a Liberal market economy; Protestant in a Coordinated market economy; Unemployed and Protestant in a Liberal market economy; Unemployed and Protestant in a Coordinated market economy; Living in a society that is Protestant and a Liberal market economy; Living in a society that is Protestant and a Coordinated market economy; Unemployed in a society that is Protestant and a Liberal market economy; Unemployed and in a society that is Protestant and a Coordinated market economy).

Table A.6.5b: Well-Being, Unemployment, and Protestantism in the Welfare State (E-A).

	Model A12	Model A13	Model A14	Model A15
Unemployed	-.342*** (.022)	-.339*** (.022)	-.322*** (.023)	-.291*** (.024)
Level 1				
Protestant	.052** (.021)	.054*** (.021)	.057** (.026)	.051** (.021)
Interactions				
Unemployed & Protestant	-	-	-.148*** (.060)	-
Level 2				
Protestant society	-.405 (.333)	-.404 (.333)	-.407 (.333)	-.749* (.413)
Liberal welfare state	1.17** (.453)	1.16** (.453)	1.14** (.453)	.978 (.636)
Conservative welfare state	1.16*** (.410)	1.17*** (.410)	1.17*** (.410)	.658 (.525)
Social democratic welfare state	1.72*** (.615)	1.72*** (.615)	1.77*** (.618)	2.05*** (.648)
Cross-level interactions				
Unemployed in a Protestant society	-	-	-	-.352*** (.058)
Unemployed in a Liberal welfare state	-	.242** (.109)	.232* (.128)	.217* (.125)
Unemployed in a Conservative welfare state	-	-.338*** (.118)	-.478*** (.146)	-.524* (.270)
Unemployed in a Social democratic welfare state	-	-.172 (.203)	-.105 (.229)	.131 (.209)
Other interactions	-	-	Included	Included
-2Loglikelihood	655,059.4	655,045.1	655,033.9	655,005.0

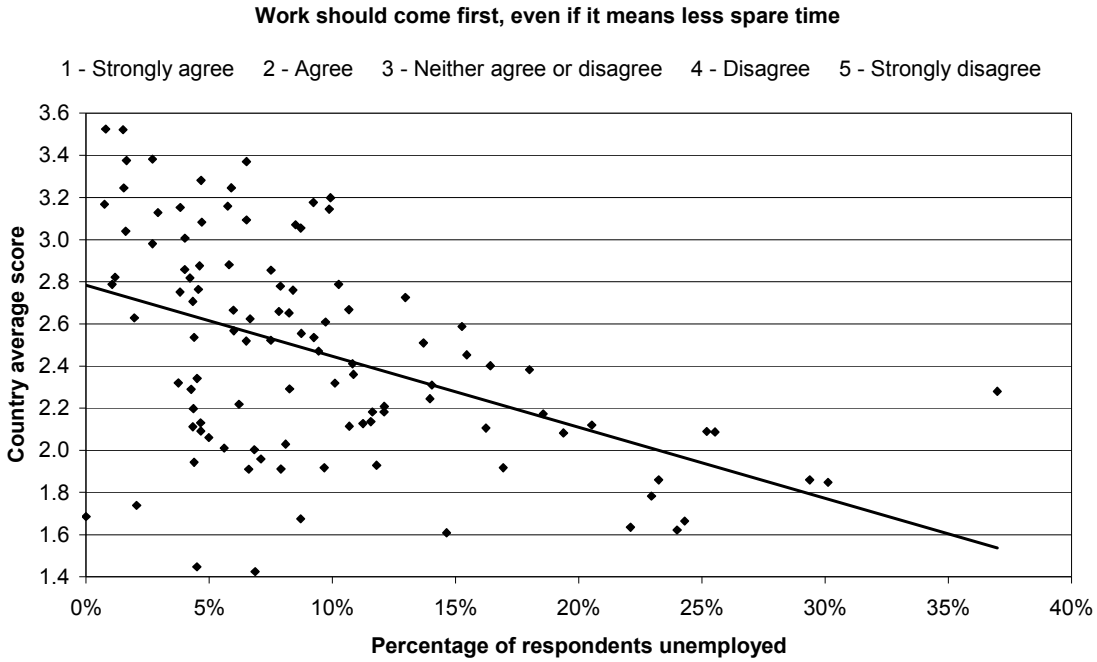
Notes: See Table 6.3. So as not to clutter the table we do not report results for the other interactions terms (Protestant in a Liberal welfare state; Protestant in a Conservative welfare state; Protestant in a Social democratic welfare state; Unemployed and Protestant in a Liberal welfare state; Unemployed and Protestant in a Conservative welfare state; Unemployed in a society that is Protestant and a Social democratic welfare state; Living in a society that is Protestant and a Liberal welfare state; Living in a society that is Protestant and a Conservative welfare state; Living in a society that is Protestant and a Social democratic welfare state; Unemployed in a society that is Protestant and a Liberal welfare state; Unemployed in a society that is Protestant and a Conservative welfare state; Unemployed in a society that is Protestant and a Social democratic welfare state).

Table 6.A.6: The Role of Religious Upbringing.

	Model A16	Model A17	Model A18	Model A19
Unemployed	-.405*** (.038)	-.336*** (.067)	-.279*** (.073)	-.261*** (.076)
Level 1				
Protestant	.117 (.072)	.042 (.035)	.140** (.059)	.037 (.035)
Raised religiously	.142*** (.030)	.134*** (.026)	.151*** (.029)	.143*** (.031)
Interactions				
Unemployed & Protestant	-	-	-.364*** (.184)	-
Unemployed & Raised religiously	-	-.096 (.077)	-.073 (.084)	-.028 (.087)
Unemployed & Raised religiously & Protestant	-	-	-.104 (.212)	-
Raised religiously & Protestant	-.098 (.078)	-	-.081 (.061)	-
Level 2				
Protestant society	.704* (.394)	.721* (.392)	.712* (.391)	.782* (.393)
Cross-level interactions				
Unemployed in a Protestant society	-	-	-	-.343** (.160)
Unemployed & raised religiously in a Protestant society	-	-	-	-.274 (.186)
Raised religiously in a Protestant society	.013 (.072)	-	-	-.034 (.056)
-2Loglikelihood	215,130.1	215,130.8	215,105.4	215,085.3

Notes: See Table 6.3. Data concern 49,245 individuals from 42 societies (see Table A.6.2).

Figure 6.A.1, Work versus Spare Time and Unemployment.



Notes: See Table 6.A.1. Country-year averages based on 137,274 individuals. The sample consists of 80 countries and 106 year observations.

Chapter 7

How United is Germany? Communist Values versus Economic Performance in East and West Germany After 1991[•]

[•] This chapter draws on joint work with Robbert Maseland, published as van Hoorn, A., and R. Maseland. 2010. Cultural Differences Between East and West Germany After 1991: Communist Values versus Economic Performance? *Journal of Economic Behavior & Organization*, 76 (3): 791-804 (van Hoorn and Maseland 2010). Helpful comments by participants of the 2008 International Socio-Economic Panel User Conference, the International Association for Research in Economic Psychology & Society for Advancement of Behavioral Economics 2008 World Meeting, and the 2009 International Society for Institutional Economics Annual Conference are gratefully acknowledged.

7.1. INTRODUCTION

The view that values and informal institutions are a cause of differences in economic performance between countries and regions is growing in popularity (Barro and McCleary 2003; Franke et al. 1991; Granato et al. 1996; Guiso et al. 2006; Harrison 1992; Jackman and Miller 1996; McCleary and Barro 2006; Noland 2005; North 1990; Pryor 2005; Swank 1996; Williamson 2000). Reunified Germany offers an excellent test case for this argument. On the one hand, the separation and, especially, the reunification of East and West Germany constitutes a large scale natural experiment giving social scientists a uniquely controlled setting to study the role of values and culture in socioeconomic developments (cf. Chapter 1). More substantively interesting is that almost two decades after political and formal institutional reunification, economic performance in East Germany is still considerably below West German standards (Boltho et al. 1997; Sinn and Westermann 2001; Sinn 2002; Hall and Ludwig 2006; Snower and Merkl 2006; Uhlig 2006, 2008). The most recent data show that in 2009 labor productivity in East Germany as a whole still stood at less than 80% of the German average (Statistisches Ämter des Bundes und der Länder 2010), while the unemployment rate in the East is 13%, almost double that in the West (Federal Statistical Office Germany 2010). An oft-heard explanation for this continuing gap is that a so-called "wall in the head" or "Mauer im Kopf" (see, for example, Corbett 2004; Häder and Häder 1995; Wagner 1999) still separates East and West Germans: persistent differences in the values held by Easterners and Westerners are thought to create divergent economic outcomes.

In this chapter, we investigate whether four decades of separation have indeed resulted in structural differences in values that are capable of causing the observed economic divergence. As with Chapter 6, part of our inspiration derives from the fact that findings concerning East-West differences in preferences thus far have been ambiguous. We apply the EP approach and estimate experienced utility functions that include a variety of (possible) determinants of SWB, including income, type of employment, and occupational status. If there are important differences in value preferences between East and West, we expect these functions to exhibit substantial heterogeneity.

Data come from the German Socio-Economic Panel (SOEP) for the period 1991-2006. Most respondents are sampled more than once, so that we apply multilevel modeling and simultaneously account for within-subject and between-subject variance. Our results show that East Germans and West Germans indeed portray important differences in values. However, an economically debilitating Communist legacy in the form of values that are not

conducive to economic performance appears largely absent. The differences between East and West that we find are often contrary to what we would expect from the literature on values and economic development. These results indicate that any relation between values and economic performance is more complex than often understood. Economic divergence seems not simply reducible to differences in values. For the debate about the transition of formerly planned economies, this would suggest that a cultural transition is not necessarily part of the process of becoming a market economy.

This chapter is organized as follows. The next section elaborates on the relation between culture and economic performance, discussing empirical work on the influence of Communism on cultural values and the limitations of this approach to the study of preferences. Section 7.3 discuses the type of experienced utility functions that we estimate, specifically the kind of (economic) determinants of SWB that we include in the empirical model, and states the hypotheses concerning a potential Communist value legacy in terms of East-West variation in the empirical structure of SWB. In Section 7.4, we discuss the data used, our models, and the method of estimation. Section 7.5 presents the results of our applying of the EP approach. We discuss our findings and relate them to economic convergence in Germany since the reunification in Section 7.6.

7.2. MARKET VALUES AND ECONOMIC PERFORMANCE

7.2.1. Cultural Differences and Economic Performance

Following the previous chapter, we trace the literature relating values to economic performance back to Max Weber's *The Protestant Ethic and the Spirit of Capitalism* (Weber 1904/1905 [1930]). Not explicitly discussed in the previous chapter, in this seminal contribution Weber included a broad explanation for the motivation behind the entrepreneurial behavior that characterizes modern capitalism, arguing that an ascetic, rational pursuit of worldly success was what drove capitalist entrepreneurs. Since then, many authors have followed up on Weber's quest to identify the kind of attitudes required for entrepreneurship. The most famous contribution is probably the one by McClelland (1961), arguing that an attitude he called "Need for Achievement" drives entrepreneurship and economic development. McClelland's claim has been tested various times with mixed results (Frey 1984; Granato et al. 1996; Gilleard 1989; Beugelsdijk and Smeets 2008). Over time, the

list of attitudes that have been said to set entrepreneurs apart has grown longer and longer to include such values as autonomy, individualism, materialism, a propensity to social recognition and risk taking, along with the classic need for achievement (e.g. Brandstätter 1997; Brockhaus 1982; Cramer et al. 2002; Cromie 2000; Fagenson 1993; Morris et al. 1994; Spence 1985; Thomas and Mueller 2000).

Authors outside the entrepreneurship literature have also sought to relate values to successful economic development. Following a second main strand in Weber's work, work ethic has been identified as one of the main factors responsible for economic prosperity (Delacroix and Nielsen 2001; Lynn 1991; see Section 6.2).[63] Next to a preference for work, a willingness to sacrifice short-term profits for long-term growth and to invest in education have been proposed as important drivers of development, especially in East Asia (Bond 1988). Interestingly, among the authors focusing on development in Asia, there seems to be a consensus that collectivism rather than individualism is important for economic success (Harrison 1992), which goes against prevalent ideas in the entrepreneurship literature. Likewise, in the literature seeking to establish cultural underpinnings of economic development in Asia, authors only rarely identify risk-taking as an important value. Apart from these differences, most of the work in this area confirms the insights from Western-oriented studies that identify values such as work ethic, materialism, a propensity to invest in education, and social recognition as prime values for development.

7.2.2. The Communist Legacy and Economic Performance: A Wall in the Head?

Many of the above-mentioned value traits may have been discouraged under Communist rule. The idea that decades of Communist rule left a heritage in the minds of people that could hamper economic development is as old as the collapse of Communism itself. Already in May 1990, even before the formal conclusion of the German reunification in early October, Shiller et al. (1991) conducted a telephone survey of random samples of the Moscow and New York populations. They sought to examine the extent to which Soviet respondents differed from their American counterparts on economic attitudes. A year later, Shiller et al. (1992) expanded their original study, adding a comparison between East and West Germans. These early efforts to find empirical support for the idea that those living in formerly Communist

[63] Let us again emphasize that Weber himself discussed value traits underpinning the historical emergence of capitalism, but never made an explicit link to economic performance.

societies would lack the attitudes needed to be successful in market economies were discouraging for those seeking a role for attitudinal factors. No significant differences were found between the populations living in formerly Communist societies and those that grew up in market economies. Although there were profound differences in behavior, these differences could not be linked to values. Instead, situational factors, factors such as the perception of economic institutions, economic expectations, and expectations about how other people will react to one's actions, were sufficient to explain them. More recently, Brodbeck and Frese (2007) argue that East and West German cultures are very much alike and characterized by the same values and attitudes. However, other studies such as Corneo and Grüner (2002) and Alesina and Fuchs-Schündeln (2007) do find (limited) evidence for differences in attitudes. Corneo and Grüner (2002) compare self-reported preferences for redistribution and inequality in six Western countries with those of six ex-Communist countries. They find that attitudinal factors play some role in explaining preferences for redistribution, although they argue that rational self-interest is the major determinant of support for governmental reduction of income inequality. Alesina and Fuchs-Schündeln (2007) present the strongest claim yet that East and West Germans have grown to hold different values. Based on a comparison of responses to survey questions about the desired role of the state in 1997 and 2002, they conclude that significant differences in values and attitudes exist, although they are slowly disappearing.

In addition to the lack of unambiguous results, there are reasons to treat the "wall in the head" thesis with a healthy dose of skepticism. We note that the thesis has a slightly uncanny resemblance with ideas that have been popular with colonial era thinkers about development. The proposition that East Germans lack such traits as "entrepreneurial spirit," "motivation," and the willingness "to assume responsibility" (Shiller et al. 1992: 127) vaguely echo ideas and myths about lazy and irresponsible natives that have been used to justify colonial domination, amongst others (Alatas 1977) (see, also, the previous chapter). Moreover, as Shiller et al. (1991, 1992), and Corneo and Grüner (2002) argue regarding formerly Communist countries, the attitudinal explanation for differences in performance might be superficial—a closer look can reveal a rationality behind differences in behavior. This is also in line with the work showing that values survey scores are highly sensitive to changes in (economic) circumstances and contexts, even to the extent that they do not measure deep-rooted value traits but capture marginal preferences (depending on levels of satiation) instead (Clarke et al. 1999; Davis et al. 1999; Duch and Taylor 1993; see Chapter 3). All this gives ample reason to treat the "wall in the head" thesis with due caution.

7.3. EXPERIENCED UTILITY AND EAST-WEST DIFFERENCES IN ECONOMIC VALUES

For the EP approach, we find that if the German division has left people from the former German Democratic Republic (GDR) and the former Federal Republic of Germany (FRG) with different attitudes and preferences, these differences should be reflected in their respective (experienced) utility functions. In our analysis, we thus estimate the structure of SWB in terms of its determinants in order to derive insights about differences in attitudes between East and West Germans. Our reasoning is that if East Germans were to value, for instance, occupational status less, occupational status should have a smaller positive influence on SWB of East Germans than of West Germans. We have the following hypothesis:

Hypothesis 1 *If differences in value orientations exist between East and West Germany, the coefficients in happiness functions differ systematically between East and West Germans.*

We are not interested in just any values differences between East and West Germans, but specifically in differences in those values that are generally associated with economic development, however. Hence, our second hypothesis is:

Hypothesis 2 *If differences in value orientations are responsible for economic divergence between East and West Germany, East Germans should attach lower weight to those values that are seen as conducive to economic performance.*

Formally testing these hypotheses involves examining how situational factors such as being married, unemployed, or living in poor health differentially affect levels of experienced utility of Easterners and Westerners. Naturally, the focus is on economic variables such as income and employment status—if an economically debilitating Communist value legacy exists, these factors will not contribute much to the SWB of people born in East Germany.

We turn to a description of our data and empirical strategy next, before presenting our results in Section 7.5.

7.4. DATA AND METHOD

7.4.1. Description of the Data

Our data come from the German Socio-Economic Panel (SOEP or GSOEP) (see Wagner et al. 2007 for a discussion). The SOEP project follows a representative sample of persons, nested in families and households since 1984, surveying them annually. Periodically, new respondents are added to the sample, and of the 5,921 households containing 12,290 individual respondents originally included in 1984, 3,476 households containing 6,203 respondents were still in the sample in 2006. To this were added, amongst others, a sample of 2,179 households with 4,553 members from the former GDR (German Democratic Republic) in 1990. On occasion, respondents have further been asked about their place of residence before reunification, allowing us to create a dummy variable that distinguishes former GDR individuals from former FRG (Federal Republic of Germany) individuals. Since we are interested in values after reunification, in our analysis we only consider observations from the period 1991-2006. In addition, we drop all individuals with missing answers on the GDR dummy and other independent variables, when applicable.

The SOEP covers a plethora of issues such as childcare, education, economic characteristics, living situation, social participation, time allocation, and personal satisfaction. The dependent variable in our statistical analyses is life satisfaction, which we relate to data about the living conditions of respondents. This measure of SWB is given by the item in the SOEP asking an individual how satisfied with life he or she is today. Answers to this question can be given on an 11-point scale ranging from "0 – low" to "10 – high." The answer categories are in discrete steps but, as previously, we follow the SWB literature in psychology and analyze this life satisfaction variable as though it is a continuous variable, knowing that this will not affect our results much (e.g. Clark et al. 2008; Ferrer-i-Carbonell and Frijters 2004; Frey and Stutzer 2000; Stevenson and Wolfers 2008).

For our independent variables we similarly select the kind of known determinants of SWB also included in our earlier analyses of heterogeneous experienced preferences, specifically those in Chapter 6. Since we are primarily interested in the idea that East and West Germans would differ from each other in ways that have economic consequences, the focus in the empirical analysis is on variables about work and income situation. These latter variables are income (in 2000 constant Euros), actual hours worked, occupational status (International Socio-Economic Index of Occupational Status; ISEI), level of education, and

employment status (whether one is Unemployed, Self-employed, a Blue collar worker, a White collar worker, a Civil servant, Out of the labor force, or Other).[64] We include the other "usual suspects" as control variables. For the SOEP dataset these are marital status, sex (male = 1), age and age squared, and health (as measured by dummies indicating whether one has an occupational disability or has stayed in the hospital in the previous year). An important difference with the EP analysis of religious variation in the preference for work in Chapter 6 is that we examine East-West variation in preferences for all these variables (although, as mentioned, the focus is on economic determinants of SWB).

To account for diminishing marginal utility of the continuous explanatory variables (income, actual working time, and occupational status) the natural logarithm is included in our regressions.[65] Correcting for nonlinearities in this way ensures that we estimate structural weights in the happiness function and that our method does not succumb to the problem of measuring marginal preferences suffered by values surveys and previous studies of cultural differences between East and West (cf. Chapter 3 and Maseland and van Hoorn 2010a).

In the estimation of our empirical models, we make a distinction between the general population, meaning the sample as a whole, and the working population, meaning individuals who at the time of observation earned a wage income and spend a non-zero amount of hours working. This choice of samples is practical as we can only include cases with non-missing data on the independent variables when estimating our empirical models. Also, including only people belonging to the working population makes a lot of sense from the perspective of the economic divergence between East and West Germany, as it is located in the labor force (lower productivity levels and higher unemployment). For our robustness checks, we take both these samples but include only observations from individuals who spent their entire formative years (0-18) in either the GDR or the FRG.

---

[64] These are dummies created using different items from the SOEP. We classify a subject as "Out of the labor force" if in year t (s)he is neither unemployed, self-employed, nor active as a blue collar worker, a white collar worker, or a civil servant. The category "Other" are those persons to whom in year t—for whatever reasons, possibly a change in employment status during the year—multiple classifications apply, e.g. both self-employed and unemployed. This category is very small (see Table 7.1).

[65] Scores for occupational status range from 16 (e.g. farmhand) to 90 (judge) with discrete one-point increments. As done for our empirical assessment of the consistency of EP measures of preference heterogeneity across indictors of SWB in Chapter 5, we treat it as a continuous measure. We thereby follow Di Tella et al. (2010) who also take the natural logarithm.

Table 7.1: Descriptive Statistics.

Variable and description	Number of observations	Mean and standard deviation		
		All	East	West
Dependent variable				
Satisfaction with life today (0-10)	260,352	6.94 (1.80)	6.44 (1.79)	7.14 (1.77)
Independent variables				
GDR (dummy, 1 = yes)	261,244	28.9% (45.3%)	-	-
Sex (1 = male)	261,244	48.2% (50.0%)	47.3% (49.9%)	48.6% (50.0%)
Age [Years]	261,244	45.7 (17.2)	44.9 (17.1)	46.1 (17.3)
Hospital stay previous year (dummy, 1 = yes)	243,740	11.8% (32.3%)	11.7% (32.1%)	11.9% (32.4%)
Occupational disability (dummy, 1 = yes)	242,982	11.4% (31.8%)	9.0% (28.6%)	12.3% (32.9%)
Marital status in survey year	261,244			
Married [base category]		62.0% (48.5%)	60.2% (49.0%)	62.8% (48.3%)
Married but separated		1.6% (12.5%)	1.5% (12.0%)	1.6% (12.6%)
Single		23.2% (42.2%)	24.5% (43.0%)	22.6% (41.8%)
Divorced		6.8% (25.1%)	7.6% (26.5%)	6.4% (24.5%)
Widowed		6.4% (24.5%)	6.2% (24.1%)	6.5% (24.7%)
Education [ISCED-1997 Classification]	257,417			
In school		1.9% (13.6%)	2.1% (14.5%)	1.8% (13.2%)
Inadequately		3.6% (18.5%)	1.1% (10.5%)	4.6% (20.9%)
General elementary [base category]		18.6% (38.9%)	11.0% (31.2%)	21.7% (41.2%)
Middle vocational		48.3% (50.0%)	53.2% (49.9%)	46.4% (49.9%)
Vocational with university entrance exam		4.2% (20.1%)	2.5% (15.7%)	4.9% (21.6%)
Higher vocational		7.1% (25.6%)	6.9% (25.4%)	7.1% (25.7%)
Higher education		16.4% (37.0%)	23.1% (42.2%)	13.6% (34.3%)
Employment Status (at time of survey)	259,416			
Self-employed		5.5% (22.8%)	4.5% (20.8%)	5.9% (23.5%)
Blue collar worker		18.5% (38.8%)	19.8% (39.9%)	17.9% (38.3%)

Table 7.1, continued.

Variable names and description	Number of observations	Mean and standard deviation		
		All	East	West
White collar worker		26.3% (44.0%)	26.3% (44.1%)	26.2% (44.0%)
Civil servant		3.8% (19.1%)	1.7% (13.0%)	4.6% (21.0%)
Unemployed [base category]		6.5% (24.6%)	11.4% (31.8%)	4.5% (20.7%)
Other		0.5% (7.2%)	0.9% (9.4%)	0.4% (6.1%)
Out of the labor force		38.7% (48.7%)	35.0% (47.7%)	40.2% (49.0%)
Current net labor income [2000 constant Euros]	152,177	1,366 (1,154)	1,080 (708)	1,481 (1,273)
Actual weekly work time [hours]	145,194	39.0 (12.6)	41.7 (11.0)	37.9 (13.0)
Occupational status [ISEI]	145,541	44.3 (16.0)	43.5 (15.6)	44.6 (16.2)

Notes: Standard deviations in parentheses. Occupational status is measured by the International Socio-Economic Index of Occupational Status (ISEI) and is based on Ganzeboom and Treiman's (1996) recoding of individuals' ISCO88 occupational classification (see Chapter 5). Due to rounding and/or the combining of multiple SOEP items (employment status; see Note 54) percentages may not add up to 100%. Because of missing observations, marital status excludes individuals whose spouse lives in his or her native country.

The SOEP data have previously been used to assess East-West differences in value preferences by Alesina and Fuchs-Schündeln (2007). They found some relevant SP items but these were included in the survey in 1997 and 2002 only. Our data in contrast spans the entire period 1991-2006 and includes more individuals who are also sampled more often, typically six to eight times (though this depends on the specific determinants of SWB included in the empirical model). Table 7.1 gives descriptive statistics on the variables used. Details of our method follow.

7.4.2. Model and Estimation

Though perhaps not obvious, just like the WVS data used in our earlier empirical analysis the SOEP data is structured hierarchically such that there is variation in life satisfaction both within an individual (Level 1) and across individuals (Level 2). The level-1, within-subjects variance is due, amongst others, to changes in personal circumstances such as an increase or decrease in income or a change in one's health status. The level-2, between-subjects variance

in happiness, on the other hand, is due to individual-specific factors that are time-invariant, most notably one's personality (see, for example, Diener et al. 1999: 279-282 for an overview of genes and personality as determinants of SWB). Because of this multilevel structure of our data we again apply multilevel or hierarchical linear modeling (Gelman and Hill 2007; Raudenbush and Bryk 2002; Snijders and Bosker 1999).

As mentioned in the previous chapter, on top of several other advantages, multilevel modeling gives us a proper way of dealing with cross-level interactions (ibidem; see, also, Chapter 3). Our goal is to estimate differences in values between East and West using heterogeneous experienced utility functions. Accordingly, in our multilevel model the time-invariant, dummy level-2 variable indicating whether the individual originally is an East (GDR) or a West German (FRG) moderates the SWB effect of level-1 factors such as income, employment status, and hours worked.

Multilevel modeling has previously been applied to analyses of the SOEP data akin to ours. Lucas et al. (2004), for example, apply the method to study adaptation to unemployment using this panel. Multilevel modeling is used to deal with variability in SWB scores that is between subjects and variability in SWB that is within subjects and due to the individual falling unemployed. Their analysis indicates that changes in unemployment have a lasting impact on SWB but that substantial adaptation takes place so that the adverse happiness effect diminishes over time. Zimmerman and Easterlin (2006) similarly use a multilevel technique to examine adaptation to changes in marital status, finding that, in the long run, the formation of marital or cohabiting unions raises SWB while their dissolution lowers it.

Spelling out the formal model at length, we have an individual j (Level 2) who is observed in year t (Level 1). LS_{tj} denotes the self-reported satisfaction with life of individual j at year t. GDR_j is a time-invariant dummy indicating whether a person is a West German (0) or an East German (1). This level-2 variable can have both a direct effect on SWB, which is of no concern to us, and a moderating effect. The size of the moderating effect, which sheds light on a Communist value legacy, is captured by the interaction term $GDR_j \times x_{tj}$, where x_{tj} stands for all possible time-varying, level-1 explanatory variables. Typically, x_{tj} comprises such personal characteristics as marital status and health but also factors that are more work-oriented such as income and occupational status. Including only the variables relevant for the analysis of East-West differences in preferences, this yields the following level-1 model (within subjects):

$$LS_{tj} = \beta_{0j} + \beta_{1j} x_{tj} + \varepsilon_{tj},$$

where β_{0j} is the intercept representing the average happiness of individual j, β_{1j} is the coefficient showing how much certain factors contribute or detract from individual j's happiness, and ε_{tj} is an error term. This is all standard. The difference comes with the level-2 modeling of the parameters of the within-subjects model. This level-2 model (between subjects) is specified as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} GDR_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} GDR_j + u_{1j},$$

and the within- and between-subjects models combine to the following overall model:

$$LS_{tj} = \gamma_{00} + \gamma_{01} GDR_j + \gamma_{10} x_{tj} + \gamma_{11} (GDR_j \times x_{tj}) + [u_{0j} + u_{1j} x_{tj} + \varepsilon_{tj}]. \qquad (7.1)$$

In this model, γ_{00} is the mean intercept across all individuals, u_{0j} is an individual-specific error term representing deviations from this mean, and γ_{01} denotes the direct SWB effect of the dummy indicating whether individual j originally is an East or a West German. The model so far is called a varying or random intercepts model that takes into account unobserved heterogeneity in SWB levels (u_{0j}), which may be due to measurement errors, genes, or other time-invariant factors. Since both γ_{00} and γ_{01} refer to differences in *levels* of SWB and not to differences in the *structure* of SWB, we pay only scant attention to them in the discussion of our results.

The specification for β_{1j} is analogous to that for β_{0j}. The parameter γ_{10} denotes the mean slope coefficient across all individuals for the level-1, time-varying independent variables included in the model such as income, employment status and hours worked (x_{tj}), whilst u_{1j} is the individual-specific random deviation from this mean slope. The model thus allows the SWB effect of different factors to vary across individuals, which is called a random slopes or varying coefficients model. For the present study, the individual-specific deviations from the mean coefficients (u_{1j}) are of no concern, and we focus only on slope heterogeneity

that is associated with being an East or a West German. This is captured by the parameter γ_{11}. If East Germans suffer a Communist value inheritance, we expect many sizable and statistically significant coefficients for the GDR interaction term. The estimate for γ_{11} thereby tells us how much more or less Easterners value certain situational factors than Westerners do. The latter's valuation of these same factors is captured by the mean slope coefficient γ_{10}. We estimate the varying-intercepts, varying-coefficients model depicted in Equation 7.1 using maximum likelihood.

7.5. EMPIRICAL RESULTS

7.5.1. Basic Results

We begin our analysis of the empirical structure of SWB among East and West Germans by estimating two baseline models, which include age, gender, and marital status as dependent variables and comprises almost all individuals in the dataset (Table 7.A.1 in the appendix). Model A1 gives the coefficient estimates for the whole sample without differentiating between East and West Germans. Results match those known from the SWB literature. The simple differentiated model (Model A2) subsequently shows that a range of factors have a varying impact on life satisfaction among East and West Germans. The models are nested so that we can use a likelihood-ratio test to assess whether allowing East-West heterogeneity in the experienced utility function improves model fit statistically significantly. This test shows that the decrease in -2Loglikelihood when moving from Model A1 to Model A2 is indeed statistically significant with $p < 0.01$ (where the value of the test statistic is 123.0 with a chi-square distribution and nine degrees of freedom). Concerning the size of the coefficients, the most noteworthy differences are that East Germans are happier outside marriage than West Germans and that they suffer less from occupational disability. Being divorced, for example, hurts Easterners 0.176 happiness points less (on the 0-10 scale) than it does Westerners. As indicated in the table, this difference is statistically significant at $p < 0.01$. Similarly, having an occupational disability lowers the happiness of East Germans by almost 0.08 points less than it does the happiness of West Germans ($p < 0.05$). This heterogeneity in the empirical structure of SWB might indicate differences in gender roles inherited from the Communist era, when social policy stimulated economic independence of women and facilitated

separation and divorce (Kolinsky and Nickel 2003). Also, these findings may reflect stronger structures of community support. Under Communist rule, East Germany is usually considered to have skipped the individualistic trend experienced by the West from the late 1960s onwards (Brodbeck and Frese 2007). In addition, religion could be a factor in these East-West differences in the SWB effect of marital status, where a lower level of religiosity of East Germans reduces the social stigma associated with divorce. As expected, there is also a substantial happiness gap between East and West (the gap is roughly 0.6; see also Table 7.1) but it again deserves emphasizing that for this paper we are only interested in heterogeneity in the structure of SWB in terms of its determinants and not in differences in levels. In the remainder, we no longer report level differences (nor intercepts).

More relevant for our purpose, further analysis reveals that variables relating to work, income and education also have different impacts on life satisfaction in East and West Germany. Table 7.2a shows the main results of a model for the general population (not including income and other job-related variables such as actual hours worked). The results in this table, specifically for Model 2, which is the differentiated model, can be taken to lend some support to the (cruder versions of the) "wall in the head" thesis. East Germans appear significantly happier to be civil servants, which adds about 0.28 to their SWB (on top of the roughly 0.86 points it adds to the happiness of West Germans). This may, with some effort, be seen to fit prejudices about lack of motivation and initiative among people growing up in a Communist system. However, it is also shown that Easterners evaluate being a blue collar worker, being a white collar worker, and being self-employed more positively (though the latter not statistically significantly so). By elimination, the only possible conclusion is that, for East Germans, not belonging to the working population is more hurtful than for West Germans. In other words, it appears that regardless of the specific type of employment East Germans have a stronger preference for working (relative to being unemployed, which is the base category). Although this result contradicts the "wall in the head" thesis, it intuitively makes sense. A possible interpretation of this is that it reflects remnants of a work ideology instilled by Communism. Alternatively, the stronger East German work ethic may be thought to represent structural differences in development between the East and West; a focus on work and material security is generally considered to diminish alongside rising incomes and levels of economic development, the so-called shift from materialist to postmaterialist values (Inglehart 1990, 1997). Less surprising are the differences in the SWB effects of education. Having enjoyed "only" vocational training (Middle vocational, Vocational with university entrance exam, or Higher vocational) is perceived much more positively among East Germans

than among West Germans. This effect of having a degree from a career or trade school corresponds with the idea that Communist societies have high valuation of manual labor and craftsmanship. The negative coefficient for the higher education interaction term is also consistent with this, though it is small (-0.019 happiness points) and not statistically significant at usual levels.

Table 7.2a: Experienced Utility Functions and East-West Differences in Preferences, General Population.

	Model 1	Model 2
In school	.275***	.229***
	(.029)	(.035)
In school * GDR	-	.154***
		(.059)
Inadequately	-.057**	-.075***
	(.025)	(.027)
Inadequately * GDR	-	.117
		(.075)
Middle vocational	.064***	.054***
	(.013)	(.015)
Middle vocational * GDR	-	.054*
		(.031)
Vocational with university entrance exam	.087***	.061**
	(.025)	(.028)
Vocational with university entrance exam * GDR	-	.148**
		(.065)
Higher vocational	.137***	.117***
	(.022)	(.025)
Higher vocational * GDR	-	.096*
		(.050)
Higher education	.321***	.331***
	(.020)	(.024)
Higher education * GDR	-	-.019
		(.046)
Self-employed	.769***	.746***
	(.022)	(.027)
Self-employed * GDR	-	.032
		(.049)
Blue collar worker	.725***	.706***
	(.015)	(.020)
Blue collar worker * GDR	-	.027
		(.031)
White collar worker	.784***	.738***
	(.015)	(.020)
White collar worker * GDR	-	.125***
		(.031)
Civil servant	.924***	.856***
	(.030)	(.034)

Table 7.2a, continued.

	Model 1	Model 2
Civil servant * GDR	-	.284*** (.076)
Other employment	.196*** (.042)	.251*** (.057)
Other employment * GDR	-	-.184** (.082)
Outside labor force	.763*** (.015)	.732*** (.019)
Outside labor force * GDR	-	.071** (.029)
-2Loglikelihood	842,820.1	842,772.5

Notes: Standard errors in parentheses. * (**) (***) indicates significance at the 10% (5%) (1%) level. All models include individual fixed effects in the form of varying intercepts. Excluded categories are "Unemployed" for Employment status and "General elementary" for Education. Controls are sex, age, age^2, marital status, hospital stay, occupational disability, and the GDR dummy (see Table 7.A.1 in the appendix). Base category is a married West German woman without a job and with general elementary education, who has not had a hospital stay last year and does not suffer an occupational disability. Estimates are based on 32,239 between-subject and 236,828 within-subject observations.

Although these findings are interesting, it is likely that these results in part pick up the income effects of work and of education. For this reason, Table 7.2b provides the same analysis but with inclusion of income and other job-related effects. As mentioned above, the sample here is limited to individuals who, at the time of observation, belonged to the working population. To keep sample size as large as possible, we have dropped the education variable for this table. Nevertheless, it is unavoidable that by adding job-related variables the number of between-subject observations is reduced by more than one-third and the number of within-subject observations by almost one-half.

Table 7.2b: Experienced Utility Functions and East-West Differences in Preferences, Working Population.

	Model 3	Model 4
Monthly net labor market income (Natural logarithm)	.303*** (.012)	.263*** (.014)
Monthly net labor market income * GDR	-	.154*** (.026)
Average actual work hours per week (Natural logarithm)	-.208*** (.014)	-.183*** (.016)
Average actual work hours per week * GDR	-	-.077** (.034)
Occupational status (Natural logarithm)	.142*** (.019)	.116*** (.022)

Table 7.2b, continued.

	Model 1	Model 2
Occupational status * GDR	-	.103** (.042)
Self-employed	.660*** (.181)	.308 (.266)
Self-employed * GDR	-	.471 (.299)
Blue collar worker	.546*** (.180)	.190 (.265)
Blue collar worker * GDR	-	.472 (.296)
White collar worker	.664*** (.180)	.304 (.265)
White collar worker * GDR	-	.486 (.296)
Civil servant	.811*** (.182)	.467* (.267)
Civil servant * GDR	-	.467 (.305)
Other employment	.089 (.185)	-.228 (.272)
Other employment * GDR	-	-.023 (.096)
-2Loglikelihood	410,101.9	410,041.6

Notes: See Table 7.2a. Base category is a married West German woman without a job, who has not had a hospital stay last year and does not suffer an occupational disability (see Table 7.A.1 in the appendix). Estimates are based on 21,118 between-subject and 120,111 within-subject observations.

7.5.2. Robustness

Our robustness checks examine the possible role of two factors. The first of these is socialization; the second rather brief check concerns a possible multicollinearity problem due to the overlap between occupational status and the type of employment (self-employed, blue collar, white collar, or civil servant), both of which involve a classification of the individual's occupation. The proposition that people originating from East and West Germany would have different values and conceptions is informed by the idea that society has an impact on the way people think and behave. We know from the values literature (e.g. Inglehart 1990, 1997) that the impact of this socialization is strongest in people's formative years. Hence, we might expect that any differences in internalized values between East and West Germans are likely to be most profound among those groups that have spent their entire formative years in either

the GDR or the FRG. Table 7.A.2 in the appendix gives the results for our baseline model, limiting the sample to individuals born between 1946 and 1971. Choosing this birth cohort ensures that we only include individuals born during Communist reign and who had reached adulthood by the time of the reunification. As in the baseline model that uses the whole sample (Table 7.A.1. in the appendix), we observe basic differences among those originating from the GDR and those having grown up in the FRG. Compared to the homogeneous model, introducing East-West heterogeneity leads to statistically significantly improved model fit (likelihood-ratio test of Model B2 nested in Model B1; $p < 0.01$). The loss of happiness caused by experiencing a divorce, for example, is decidedly lower in East Germany than in the West (-0.311 versus -0.181 happiness points). This is again an indication that gender roles and informal systems of communal support might be functioning differently in the so-called new states compared to the old states.

If we move to the effects of work-related variables (Table 7.3a), we find that limiting ourselves to the subsample of people born between 1946 and 1971 does not substantially alter the results (cf. Table 7.2a). Again, Easterners consistently value having a job higher than Westerners do, and this holds independent of type of employment. With regard to education, the more positive evaluation of having higher vocational training as highest education remains (plus 0.048 for Middle vocational, plus 0.205 for Vocational with university entrance exam, and plus 0.071 for Higher vocational), though the estimates for Model 6 are somewhat less precise than for Model 2 (which is expected given the smaller sample size). At the same time, not having finished any level of schooling appears much more problematic for East Germans than for West Germans, which previously it did not (-0.609 in Model 6 versus 0.117 in Model 2). This may be related to the fact that only very few East Germans born between 1946 and 1971 have obtained an inadequate level of education.

Table 7.3a: Experienced Utility Functions and East-West Differences in Preferences, General Population Born 1946-1971.

	Model 5	Model 6
In school	1.16*** (.341)	1.16*** (.341)
In school * GDR	-	
Inadequately	-.125*** (.041)	-.111*** (.042)
Inadequately * GDR	-	-.609*** (.218)

Table 7.3a, continued.

	Model 5	Model 6
Middle vocational	.112***	.108***
	(.021)	(.022)
Middle vocational * GDR	-	.048
		(.057)
Vocational with university entrance exam	.147***	.125***
	(.035)	(.038)
Vocational with university entrance exam * GDR	-	.205*
		(.107)
Higher vocational	.160***	.150***
	(.030)	(.033)
Higher vocational * GDR	-	.071
		(.078)
Higher education	.363***	.349***
	(.028)	(.032)
Higher education * GDR	-	.080
		(.072)
Self-employed	.885***	.870***
	(.027)	(.034)
Self-employed * GDR	-	.049
		(.059)
Blue collar worker	.849***	.835***
	(.020)	(.026)
Blue collar worker * GDR	-	.045
		(.040)
White collar worker	.941***	.916***
	(.020)	(.026)
White collar worker * GDR	-	.094**
		(.041)
Civil servant	1.16***	1.11***
	(.038)	(.043)
Civil servant * GDR	-	.327***
		(.097)
Other employment	.293***	.397***
	(.050)	(.070)
Other employment * GDR	-	-.271***
		(.098)
Outside labor force	.759***	.785***
	(.021)	(.027)
Outside labor force * GDR	-	-.184***
		(.047)
-2Loglikelihood	422,478.8	422,400.8

Notes: See Table 7.2a. Sample is limited to individuals born after 1945 and before 1972. Estimates are based on 15,028 between-subject and 120,658 within-subject observations. There are no East Germans born between 1946 and 1971 still in school.

Limiting the working population sample by our birth year criterion does not substantially alter results either (Table 7.3b). Both income and status remain more important

in the former GDR—the latter even more than doubly so, whereas Easterners also seem to value leisure more highly, though not statistically significantly so. The most striking result of this model is the higher valuation of self-employment among those having spent their formative years in the GDR. Next to East Germans having a stronger preference for working in general, which is already contrary to what the "wall in the head" thesis would predict, this would suggest West Germans may actually lack in entrepreneurial spirit.

Table 7.3b: Experienced Utility Functions and East-West Differences in Preferences, Working Population Born 1946-1971.

	Model 7	Model 8
Monthly net labor market income (Natural logarithm)	.340***	.291***
	(.014)	(.017)
Monthly net labor market income * GDR	-	.200***
		(.032)
Average actual work hours per week (Natural logarithm)	-.188***	-.169***
	(.018)	(.020)
Average actual work hours per week * GDR	-	-.015
		(.042)
Occupational status (Natural logarithm)	.123***	.092***
	(.022)	(.026)
Occupational status * GDR	-	.126**
		(.050)
Self-employed	.796***	.256
	(.210)	(.324)
Self-employed * GDR	-	.706**
		(.356)
Blue collar worker	.700***	.156
	(.209)	(.323)
Blue collar worker * GDR	-	.696**
		(.353)
White collar worker	.825***	.292
	(.209)	(.323)
White collar worker * GDR	-	.659*
		(.353)
Civil servant	.996***	.462
	(.212)	(.325)
Civil servant * GDR	-	.711**
		(.363)
Other employment	.298	-.194
	(.215)	(.331)
Other employment * GDR	-	.035
		(.117)
-2Loglikelihood	299,369.8	299,298.2

Notes: See Table 7.3a. Estimates are based on 13,389 between-subject and 88,069 within-subject observations.

For the second robustness check, we examine whether our results for the (differential) SWB effects of occupational status and the different types of employment are sensitive to a possible multicollinearity problem resulting from the correlation between these two factors. Results depicted in Table 7.A.3 in the appendix show they are not. Relative to Model 4 (Table 7.2b) the statistical fit of the models decreases significantly as it logically should. Quantitatively we find that occupational status becomes a more important determinant of happiness with type of employment excluded, but that East Germans still value status much more than West Germans do (0.279 versus 0.190). Similarly, the East-West difference in valuation of different types of employment remains roughly the same (Model 4 versus Model C2).

Overall, we conclude that our findings on East-West differences in preferences, specifically concerning market or economic values, are robust. They are not sensitive to the sample of individuals included and do not depend on any particular model specification.

7.6. DISCUSSION AND CONCLUSION

We have scrutinized the thesis that four decades of division in Germany has caused an enduring legacy in terms of differences in values and attitudes. This so-called "wall in the head" or "Mauer im Kopf" is, in turn, thought to account for the continuing gap in economic performance between East and West. According to this thesis, East Germans lack the traits required for successfully operating in a market economy, such as the valuation of autonomy and materialism, a propensity to social recognition, risk taking and assuming responsibility, and a felt need for achievement. To investigate this proposition, we apply the EP approach. Specifically, we estimate and then compare experienced utility functions for East and West German respondents, using SWB scores as a proxy for experienced utility.

Results show that for the period after reunification until 2006, there are significant differences in preferences between East and West Germans. East German culture turns out to be more favorable to being divorced or widowed, for example, which may reflect different gender, family and community roles between East and West. Also, with respect to economic values such as a materialism and preferences for employment, East Germans differ from West Germans in ways that may be seen to reflect values instilled during Communism.

Although the EP method is able to uncover important differences between East and West, these differences generally do not correspond to what proponents of the "wall in the

head" thesis might expect. East German values do not stand out as less compatible with economic success in a market economy. Some do: East Germans appear to have a slightly greater dislike of hours spent working and seem to have a lower esteem for academic education relative to vocational education (associated with specific occupations), both of which may have negative effects on income and growth levels. The more fundamental differences in experienced preferences between East and West Germans go against the "wall in the head" thesis though. Rather than being less motivated by income and occupational status, East Germans turn out to attach much more importance to these factors than West Germans do. In addition, they appear to have a substantially higher preference for working (as opposed to being unemployed). All this does not indicate a lack of the motivation needed to perform well in a market economy. Even more at odds with the "wall in head" thesis, there is also some evidence that East Germans value being self-employed higher than West Germans do. We therefore conclude that a Communist inheritance in the form of less market-oriented values seems to be lacking. Although there are differences, East German values appear to bear little relation to preconceptions about East Germans underlying the "wall in the head" thesis. If anything, East Germans appear to have values that in many respects are more rather than less compatible with a market economy.

Alternative interpretations of our findings are also possible. One may argue, for instance, that the greater value East Germans attach to income and occupational status are due to a reference-group effect for which Easterners mainly compare themselves to Westerners who, on average, enjoy both higher income and higher status. Such an account suggests that value differences can also have non-cultural sources and partially derive from social comparison. These possibilities do not change the result that East Germans attach greater value to economic circumstances, however, let alone that they would provide any support for a values-based explanation of the persistent East-West gap in economic performance. Regardless of the origins of their preferences, the differences between East and West Germans that we observe do not portray a pattern that may be usefully related to economic divergence. All in all, it seems that the idea that differences in economic preferences and attitudes are behind the persistent differences in economic performance between East and West Germany does not have much basis. East is East and West is West, but the twain meet in unexpected ways.

APPENDIX TO CHAPTER 7

Table 7.A.1: Baseline Model of Heterogeneous Experienced Utility Functions.

	Model A1	Model A2
Intercept	8.40***	8.36***
	(.048)	(.055)
GDR dummy	-.692***	-.501***
	(.017)	(.108)
Sex (Male=1)	-.007	-.001
	(.015)	(.017)
Sex * GDR	-	-.021
		(.034)
Age	-3.37***	-3.05***
	(.190)	(.221)
Age * GDR	-	-1.40***
		(.433)
Age^2	2.15***	1.76***
	(.188)	(.219)
Age^2 * GDR	-	1.66***
		(.432)
Married but separated	-.570***	-.630***
	(.026)	(.031)
Married but separated * GDR	-	.222***
		(.058)
Single	-.231***	-.245***
	(.017)	(.019)
Single * GDR	-	.057
		(.039)
Divorced	-.314***	-.364***
	(.019)	(.023)
Divorced * GDR	-	.176***
		(.042)
Widowed	-.333***	-.430***
	(.024)	(.029)
Widowed * GDR	-	.335***
		(.054)
Hospital stay (0 = No)	-.220***	-.211***
	(.009)	(.011)
Hospital stay * GDR	-	-.036*
		(.020)
Occupational disability (0 = No)	-.493***	-.511***
	(.014)	(.017)
Occupational disability * GDR	-	.079**
		(.034)
-2Loglikelihood	865,777.8	865,654.8

Notes: Standard errors in parentheses. * (**) (***) indicates significance at the 10% (5%) (1%) level. All models include individual fixed effects in the form of varying intercepts. Excluded category for marital status is "Married" so that the base category is a married West German woman, who has not had a hospital stay last year and does not suffer an occupational disability. For scaling purposes, age is divided by 100 and age squared by 10,000. Estimates are based on 32,362 between-subject and 241,891 within-subject observations.

Table 7.A.2: Robustness of Baseline Model, Sample Born 1946-1971.

	Model B1	Model B2
Sex (Male=1)	-.044**	-.063**
	(.021)	(.025)
Sex * GDR	-	.072
		(.049)
Age	-3.18***	-4.37***
	(.531)	(.605)
Age * GDR	-	5.36***
		(1.26)
Age^2	.541	1.79**
	(.644)	(.736)
Age^2 * GDR	-	-5.73***
		(1.52)
Married but separated	-.552***	-.618***
	(.029)	(.0350
Married but separated * GDR	-	.220***
		(.064)
Single	-.232***	-.246***
	(.021)	(.024)
Single * GDR	-	.050
		(.052)
Divorced	-.271***	-.311***
	(.022)	(.026)
Divorced * GDR	-	.130***
		(.048)
Widowed	-.447***	-.679***
	(.060)	(.074)
Widowed * GDR	-	.672***
		(.125)
Hospital stay (0 = No)	-.163***	-.150***
	(.013)	(.015)
Hospital stay * GDR	-	-.052*
		(.030)
Occupational disability (0 = No)	-.505***	-.517***
	(.024)	(.028)
Occupational disability * GDR	-	.050
		(.056)
-2Loglikelihood	432,796.0	432,715.7

Notes: See Table 7.A.1. Sample is limited to individuals born after 1945 and before 1972. Estimates are based on 15,080 between-subject and 122,633 within-subject observations.

Table 7.A.3: Robustness of East-West Differences in the Evaluation of Occupational Status and Type of Employment.

	Model C1	Model C2
Monthly net labor market income (Natural logarithm)	.285***	.277***
	(.014)	(.013)
Monthly net labor market income * GDR	.176***	.163***
	(.025)	(.026)
Average actual work hours per week (Natural logarithm)	-.177***	-.184***
	(.016)	(.016)
Average actual work hours per week * GDR	-.046	-.078**
	(.033)	(.034)
Occupational status (Natural logarithm)	.190***	-
	(.020)	
Occupational status * GDR	.089**	-
	(.039)	
Self-employed	-	.317
		(.266)
Self-employed * GDR	-	.467
		(.299)
Blue collar worker	-	.163
		(.265)
Blue collar worker * GDR	-	.449
		(.296)
White collar worker	-	.312
		(.265)
White collar worker * GDR	-	.489*
		(.296)
Civil servant	-	.495*
		(.267)
Civil servant * GDR	-	.480
		(.304)
Other employment	-	-.239
		(.272)
Other employment * GDR	-	-.019
		(.096)
-2Loglikelihood	410,254.6	410,106.8

Notes: See Table 7.2b.

Summary and Conclusion of Part III

The two chapters in the third part of this thesis have applied the EP approach to shed new light on two long-standing issues in the study of the economics of preference heterogeneity. Over a century ago, Max Weber (1904/1905 [1930]) hypothesized a specific Protestant work ethic. Since then, many researchers have sought to investigate his thesis empirically and over the last couple of years several economists have followed suit, examining religious patterns in the preference for work and other work-related values. In fact, in the wake of 9-11 interest in religion as a source of differences in sociopolitical and socioeconomic structures seems to experience a revival. At the same time, results on Weber's original thesis have been counterintuitive, ambiguous and overall confusing, a century of research notwithstanding. Chapter 6 set out to test whether a specific Protestant work ethic exists. Applying the EP method, we operationalize work ethic intuitively as the psychic cost of joblessness as reflected in unemployment's adverse effect on individuals' experienced utility. The rationale is that higher work ethic implies that people are hurt more by not having a job. Because the most prominent multi-country questionnaires have included items concerning both SWB and respondents' employment status, we are able to give the most comprehensive assessment of Weber's thesis thus far, covering almost 150,000 individuals from 82 societies. We find empirical evidence on a Protestant work ethic at two levels. First, unemployment hurts Protestants more, and, second, the psychic cost of unemployment is greater in Protestant societies. Further analysis shows these results are robust, and that the effects are independent of social stigma effects as with a moderating effect of peer unemployment. Over a century after Weber first posited his thesis, we are able to show that Protestantism is associated with greater value attached to work.

Along similar lines, Chapter 7 examines differences in value preferences between East and West Germany as a source of the substantial gap in economic performance that remains between East and West Germany even 20 years after formal reunification. Since formal institutions have been equalized, East and West Germany serve as a one-of-a-kind natural experiment for economists and other social scientists. Specifically, it offers a unique environment for testing the role of culture and other informal institutions in economic development. In fact, many researchers have argued that East Germans suffer a "wall in the head" ("Mauer im Kopf"): an economically debilitating value depositions instilled in them by 40 years of Communist reign. Previous empirical analyses of this thesis produced decidedly

mixed results. We apply the EP approach and revisit this issue. In a panel consisting of 15,000 to more than 30,000 individuals, on average sampled between six to eight times, we compare the shape and content of the experienced utility function for East and West Germans. Our focus thereby is on preferences with a likely effect on economic performance, notably the value attached to work, income, and occupational status. Results show that preferences indeed vary between East and West Germans. The nature of the differences we find is not the kind that one would expect from a cultural explanation for the East-West gap in economic performance, however. If anything, East Germans entertain values more conducive to economic performance. The chapter therefore concludes that East Germans suffering a Communist cultural legacy appears largely a myth.

Together, the empirical applications in Chapters 6 and 7 demonstrate the ability of the EP approach to make significant contributions to the emerging economics of preference heterogeneity field. A particularly noteworthy feature of both applications subsequently is that they are also a showcase for the flexibility of the EP approach. Earlier assessments of a specific Protestant work ethic (e.g. Norris and Inglehart 2004; Schaltegger and Torgler 2010; Torgler and Schaltegger 2009) and East-West differences in preferences (e.g. Corneo and Grüner 2002; Alesina and Fuchs-Schündeln 2007) typically had to draw on items available from surveys more or less by chance. The EP approach tends to make less strong demands on the availability of suitable survey data. As discussed in Chapter 2, experienced preferences can be measured for many states of affairs; all that is required is that measures of experienced utility can be matched to relevant determinants of SWB. Accordingly, the empirical analyses presented in Chapters 6 and 7 have been most inclusive, covering more societies and more individuals than earlier work, and/or have been able to cover group heterogeneity in preferences over a much broader range of issues. In the conclusion of this thesis we include a discussion of a most promising avenue for future research applying the EP method.

Part IV: Conclusion

Chapter 8

Conclusion: Taking Stock and Looking Ahead

8.1. THIS THESIS

This thesis has been on the measurement of preferences for the purpose of studying the economics of preference heterogeneity. Economists have long shunned the study of preferences—and for good reason. A watershed appears to have taken place, however, and economists are increasingly debating and empirically scrutinizing the shape and content of the utility function.

The three preceding parts constitute the heart of this thesis. The first two of these deal mostly with issues of method. Measurement—giving empirical content to the construct we seek to investigate in a way that renders reliable and valid measures—is essential for the study of preferences; this is the key message in Stigler and Becker's (1977) famous de gustibus paper and its plea for taking preferences as identical and constant. The simple but fundamental insight contained in their argument is that we need preference heterogeneity to be observable before it can be meaningfully invoked in the study of economic phenomena. In Part I (Chapters 2 and 3) I reviewed the three main approaches to measuring preferences and assessing preference heterogeneity. Underlying this discussion of the quantification of preferences (and the thesis as a whole) is an intuitive understanding of preferences as tastes, which involves mental entities including people's attitudes and objectives but also the criteria by which they evaluate outcomes or different states of affairs. The three approaches discerned are revealed preferences (RP), stated preferences (SP), and what can best be dubbed experienced preferences (EP). The latter involves using data on subjective well-being to estimate heterogeneous experienced utility functions. These three methods are the same methods that are applied in such valuation exercises as calculating compensating schedules for noise nuisance and air pollution or in the construction of money-metric welfare indicators. Next to introducing the approaches, the main goal of the review was to evaluate the usefulness of the approaches from the perspective of their application in the study of the economics of preference heterogeneity. The thesis uncovered some important limitations with the usability of the RP approach (i.e. the potential of applying RP measures in the study of the economics of preference heterogeneity). In short, RP measures of preference heterogeneity are not widely available and appear ill-suited for explanation as they do not allow clear separation of the explanans (the thing that explains) and the explanandum (the thing that requires an explanation). Similarly, the SP approach appeared prone to a major measurement problem, resulting in invalid measures of preference heterogeneity, diminishing

the usefulness of this type of measures of heterogeneous preferences in the study of the economics of preference heterogeneity.

The thesis identified the EP approach as having a great deal of potential, though this was largely undemonstrated. Complementing the evaluation in Part I, Part II of the thesis (Chapters 4 and 5) contains a detailed assessment of the EP approach, starting with an elaboration of the empirical subjective well-being construct (commonly abbreviated as SWB) that provides the essential input for empirically estimating (heterogeneous) experienced utility functions. Following my initial discussion of SWB—how it is defined, how it is measured, and how it can be applied—I examined whether EP measures of preference heterogeneity are reliable and valid. I did so in two steps. First I surveyed evidence on the psychometric quality of SWB indicators. Second, I empirically analyzed whether the EP approach renders meaningful measures of preference heterogeneity. The psychometric assessment of EP measures in Part II provides the foundation for applying the EP approach to issues of substantive interest. Part III of the thesis (Chapters 6 and 7) has studied two such issues. The first question this thesis sought to answer is whether there is a Protestant work ethic, as hypothesized by Max Weber in his landmark study of *The Protestant Ethic and the Spirit of Capitalism*, generally regarded as the first work looking at a role for heterogeneous preferences in social and economic phenomena. The second question concerns the gap in economic performance that remains between East and West Germany two decades after reunification. Since formal institutions have been equalized among the former states, a typical explanation is that over 40 years of Communist reign must have instilled in East German minds an economically debilitating cultural legacy. I tested whether such a "Mauer im Kopf" or "wall in the head" exists and whether the differences in preferences found may explain the differences in economic performance between the former German Democratic Republic and the former Federal Republic of Germany.

The remainder of this concluding chapter is organized as follows. Section 8.2 gives a summary of the thesis' chief empirical findings. Section 8.3 discusses some limitations, both specific to this thesis and its development of the EP method and general, concerning the whole field of the economics of preference heterogeneity. Section 8.4 concludes with a tentative outlook on the study of the economics of preference heterogeneity that also includes two concrete suggestions for future empirical work on heterogeneous preferences.

8.2. SUMMARY OF MAIN EMPIRICAL FINDINGS

This thesis' main empirical findings concern the following four relations:

(i) the negative values-practices correlations;

(ii) the consistency of preference heterogeneity measured using happiness scores and preference heterogeneity measured using satisfaction ratings;

(iii) the effect of Protestantism on the psychic cost of unemployment; and

(iv) differences between East Germans and West Germans in values conducive to economic performance.

Finding I: Negative Values-Practices Correlations

The first of the four main empirical findings concerns the values-practices relation in SP data, notably values scores from values surveys. My evaluation of the psychometric quality of SP data of (heterogeneous) preferences in Chapter 2 noted the sensitivity of such data to changes in circumstances. This chapter also mentioned the possible "marginal preferences problem:" the critique that SP measures of preference heterogeneity such as those available from the World Values Survey or the widely used Hofstede framework capture group differences in marginal preferences—the preference for increasing satiation of an objective given current levels of satiation—rather than values or attitudes—the preference for satiating the objective in general.

Chapter 3 expanded on this potential problem. A string of recent research on preferences reports counterintuitive, even paradoxical negative relations between reported values and actual practices. I find that such findings make a great deal of sense once one realizes the distinction between marginal preferences and values or attitudes at large. I subsequently demonstrated this idea by using it to account for the alleged Muslim democracy paradox: the fact that Muslim-majority countries tend to be less democratic, while both individual Muslims and individuals in Muslim-majority countries tend to have a much more positive attitude towards democracy than other people do. The empirical evidence shows that this "paradox" actually reflects the general pattern that people view democracy more favorable the less democratic their political environment.

The marginal preferences problem turns out to be pervasive. Notably, closer scrutiny of some work studying differences in work ethic reveals that these data consistently suffer from the marginal preferences problem.

Finding II: Consistency of Heterogeneity in Experienced Preferences

The second key empirical relation concerns the psychometric quality of (heterogeneous) experienced preferences. I chose to err on the safe side, viewing the reliability and validity of SWB data as a necessary but not a sufficient condition to ensure that EP measures of preference heterogeneity are meaningful. SWB data are relatively new to economics and almost no attention has been paid to the psychometric quality of SWB-based preference measures.

For this thesis, I thoroughly assessed the reliability and validity of EP measures of preference heterogeneity. My approach was to estimate experienced utility functions for a large number of countries using two different SWB indicators and then examine the extent to which heterogeneity found using one SWB indicator replicates itself in heterogeneity found using the other indicator. If EP measures are reliable and valid we expect different SWB indicators to render roughly similar patterns of country variation in experienced preferences—even though the different indicators may actually target different components of the over-arching SWB construct. At the least, meaningful EP measures require that the measured heterogeneity has some level of consistency—and not be contradictory, whilst higher levels of consistency indicate that group variation in experienced preferences is mostly of a structural rather than a random nature.

I assessed EP measures of preference heterogeneity applying this approach in Chapter 5. The assessment involved a measure of happiness, which is associated most with the affective part of SWB, and a measure of satisfaction, associated most with the cognitive part of SWB. Empirical results show that variation in experienced preferences can be quite consistent across measures of SWB and that country variation in preferences using one indicator may replicate itself in country variation in the other indicator to a large extent.

Consistency can be a result of many factors, however. I identified and examined a variety of these. The most important factor by far is the number of observations (per country). The more data are available, the higher the consistency between the happiness and satisfaction coefficients estimated separately for all countries in the sample. Another factor to have an effect, albeit small, is whether consistency is assessed for standardized or unstandardized coefficients.

Finding III: The effect of Protestantism on the Psychic Cost of Unemployment

This thesis has paid a lot of attention to the measurement of preferences and the assessment of heterogeneity therein (see, also, the first two findings summarized above). In

no small part, my extensive analysis of the measurement of preferences and the issues involved has been instrumental: these need to be dealt with carefully and comprehensively before it becomes possible to conduct a meaningful study of preferences and preference heterogeneity. Parts I and II of the thesis identified the EP method as the most useful approach to the study of the economics of preference heterogeneity. The first issue of substantive interest that we examined using the EP method is the existence of a specific Protestant work ethic. Weber's (1904/1905 [1930]) work on Protestant ethic and the advent of modern capitalism has received much empirical follow-up investigating whether Protestants have a distinct work ethic that sets them apart from individuals from other denominations (notably Catholics), and also extending upon Weber's original analysis and examining the effect of religion and work ethic on economic performance. Results thus far have been unsupportive of Weber's thesis and sometimes dismissive or plain counter-intuitive.

I have aimed to shed new light on this issue by applying the EP method to a comprehensive sample including some 150,000 individuals from 82 societies (data from the World Values Survey). The results of the empirical analysis can be summarized simply as that a Protestant work ethic exists, even today. Importantly, I find that the effect operates mostly at the societal level: both Protestants and individuals living in Protestant societies are hurt more by unemployment ceteris paribus, but the former effect of individual Protestantism disappears when the effect of living in a Protestant society is taken into account.

Finding IV: Differences Between East Germans and West Germans in Values Conducive to Economic Performance

The second application of the EP method has been to study the idea that a so-called wall in the head ("Mauer im Kopf") lies at the root of persistent economic disparities between East and West Germany even after two decades of formal reunification. I used data from the German Socio-Economic Panel (SOEP or GSOEP) and samples comprised some 15,000 to 30,000 individuals, with six to eight yearly observations available for each of them.

Results show that preferences indeed differ between East and West Germans. However, the variation found is not in line with the differences associated with the lagging economic performance of the former German Democratic Republic. Rather the opposite, East Germans appear to entertain values that are more not less conducive to economic performance than those of West Germans. Accordingly, Chapter 7 dismisses the idea that an

economically debilitating Communist cultural legacy accounts for the East-West gap in economic performance.

8.3. LIMITATIONS: GENERAL AND SPECIFIC

8.3.1. The Endogeneity of Preferences

A general and fundamental issue in the study of the economics of preference heterogeneity concerns the endogeneity of preferences: the idea that markets and other institutions not only affect the allocation of resources but also influence how values and tastes evolve (Bowles 1998: 75). The endogeneity of preferences poses a serious difficulty for the study of the economics of preference heterogeneity, mostly for work seeking to find a role for preferences in differences in economic performance. Roughly the problem is that, at any point in time, the same outcomes that one seeks to illuminate in terms of differences in preferences are a causal factor through which set differences have come about in the first place. In this thesis this problem lurks in the background, surfacing most prominently in one of the thesis' main empirical results, the negative values-practices correlations in SP data (see Chapter 3). I have taken these correlations to demonstrate the likely invalidity of SP measures of preference heterogeneity, mistaking marginal preferences for deep-rooted value dispositions. It is clear, however, that the marginal preferences problem of SP data is highly analogous to the issue of endogenous preferences: the circumstances in which people find themselves shape their preferences, which, in turn, shape their environments. In this sense, the marginality of SP measures of preference heterogeneity is part of the broader challenge of separating the effect of markets and other institutions on allocations from the effect of (heterogeneous) preferences on these allocations.

Different approaches and techniques have been applied to address the endogeneity problem in the study of the economics of preference heterogeneity (Fernández 2008, 2010). One of these is the use of historical case studies such as those by Lal (1998) and Landes (1998). Another one is the so-called epidemiological approach, which, by looking at migrants, is able to assess cultural differences between groups whilst keeping the institutional influence the same for all groups concerned. Thus far, the latter approach has only been applied in the analysis of RP measures of preference heterogeneity as in studies of cultural differences in fertility and female labor force participation (Fernández 2007; Fernández and

Fogli 2006, 2009). The epidemiological approach does not specify a particular type of preference measure, however.

Interestingly, recent developments in neuroeconomics and genoeconomics suggest a third way of dealing with the issue of endogenous preferences. Chapter 2 briefly noted work by economists (teaming up with scientists from other fields) relating differences in preferences to such biological factors as oxytocin (a neurotransmitter) and testosterone levels. Studies like these fit the wider literature that looks at the evolutionary and biological foundations of (heterogeneous) preferences. My outlook on trends in preference research in the next section considers work in this broad area, also discussing how some of its results and insights may be applied to address the endogeneity of preferences in studies of the economics of preference heterogeneity.

8.3.2. Groups and the Need for Variation when Estimating Heterogeneous Experienced Utility Functions

An important distinction between the RP and SP method on the one hand and the EP method on the other concerns the statistical procedure through which they give empirical content to group differences in preferences. For the RP and SP method obtaining group scores tends to involve aggregation, specifically of choices made or answers given by individuals belonging to different collectives. The EP approach, in contrast, uses estimation, taking all individual observations belonging to a particular class and empirically modeling the shape (and content) of the utility function for the selected class. This difference is not without implications. Most directly, the EP approach requires that within a group there is a sufficient amount of individual variation in the determinants of experienced utility. If there is only little variation in these variables, the estimation and therefore the measurement of preference heterogeneity will be inaccurate or even become impossible. In Chapter 2 I explained that this means that EP measures are bound to have lower resolution than either RP or SP measures: the smallest observational unit for which heterogeneous preferences can be measured is larger when using the EP approach than when using the RP or SP approach.

Having applied the EP approach to important questions in the economics of preference heterogeneity field, I can discuss in more detail how the need for variation in the determinants of experienced utility can limit the usability of the EP approach and how this limitation may be overcome. To illustrate, take the analysis in Chapter 6 of Protestantism (individual or societal) and how it affects the psychic cost of unemployment. The essence of

this analysis is to compare the (negative) SWB effect of not having a job relative to being full-time employed for different groups, Protestants versus non-Protestants, and/or people living in Protestant societies versus people living in non-Protestant societies. To be able to estimate this negative effect for these groups requires that each of them contains both unemployed and full-time employed individuals. If either of these are missing we cannot estimate the psychic cost of unemployment let alone compare this across the groups included in our analysis.

This need for variation within the groups for which preferences are assessed has not been problematic for the empirical analyses in this thesis; both our analysis of Protestant work ethic and of East-West differences in preferences contained enough variation in each of the groups compared to allow an accurate estimation of the experienced utility function. The need for variation may be problematic in other EP assessments of heterogeneous preferences, however. The final section proposes applying the EP method to the construction of comprehensive quantitative frameworks of cultural differences between countries and other classes along the lines of Hofstede (1980, 2001) and the GLOBE (Global Leadership and Organizational Behavior Effectiveness) project (House et al. 2004) as a most promising avenue for future research. The large scope of such a study makes it more likely that the need for within-group variation in the determinants SWB becomes a limiting factor. Hence, below I also suggest a way to circumvent this potential limitation of the EP approach to assessing preference heterogeneity. This final section follows next.

8.4. QUO VADIS?

This thesis has contributed to a specific area in preference research, namely the measurement of preferences and the documentation of important and substantial group differences therein. This is also the kind of work that most economists studying preferences have been engaged in. Looking ahead, several trends in and opportunities for the study of the economics of preference heterogeneity can be discerned. This section discusses two of these. I first put forward two concrete questions for empirical research on preference heterogeneity, whereby the focus—as with the thesis—is on the (mis)measurement of preferences and preference heterogeneity. More tentatively, I consider work on the evolutionary and biological foundations of heterogeneous preferences, which I relate to the fundamental issue of the endogeneity of preferences.

8.4.1. Future Directions for Preference Research

The economics of preference heterogeneity is a rapidly developing field and in this sense we may expect many interesting directions for future research. Here I propose two pragmatic avenues for follow-up research on preference heterogeneity, particularly its measurement. The first of these is somewhat defensive, and derives from the marginal preferences problem of SP data, expanded on and demonstrated extensively in Part I of this thesis. The second one is more positive and involves applying the EP method.

To start with the former, consider the following two observations, that (1) the GLOBE study finds negative values-practices correlations in seven out of its nine dimensions, and that (2) even if not all SP measures are equally sensitive to the marginal preferences problem, SP measures still seem to capture some mixture of two opposing constructs: values, which correlate positively with practices, and marginal preferences, which correlate negatively with practices. These two observations suggest that we likely have misinterpreted what frameworks such as those of Hofstede or GLOBE are measuring or that, at the least, it is unclear what exactly these frameworks are measuring. Accordingly, my first suggestion for future research is to re-assess what we think we know about preference heterogeneity based on SP data, including not only the two frameworks just mentioned but also the World Values Survey and other large-scale projects like it.[66]

Such datasets and, especially, these frameworks are widely used—literally in thousands of empirical studies. However, if we are not sure what they are measuring or, worse, if we have taken them as measuring something else than they are actually measuring, much of the knowledge gathered from applying them may be doubted.[67] This holds for both descriptive studies of preference heterogeneity based on them and for work that has used the frameworks to illuminate cross-country differences in socioeconomic outcomes. Mostly, this

[66] See, also, Taras et al.'s (2010) plea by which they conclude their contribution to the recent debate in the *Journal of International Business Studies* on Hofstede's framework, the GLOBE study, and the (mis)measurement of culture: "… if we only look to distant frontiers in search of new interesting questions for future cross-cultural research, we may miss intriguing research opportunities right in our own backyard. Just when we thought we had figured out culture, we come back to the very basics of culture conceptualization and measurement, and face some of the most fundamental questions again. We believe that answering these questions is where a great deal of excitement and promise lies for future cross-cultural research." (Taras et al. 2010: 1337).

[67] As American comedian Will Rogers notes, "the trouble isn't what people don't know; it's what they do know that isn't so" (Rogers as cited by Boulding 1966: 1).

proposed re-assessment would involve altering our mindset to have a different attitude towards negative empirical results or results that seem counterintuitive (as with the marginal preferences problem). Compared to earlier work, researchers may be more inclined to explore such findings and seek to identify factors that can possibly account for them, for instance by including an expanded set of explanatory variables in their empirical models.

My second proposed direction for preference research is not entirely unrelated to the first as it takes inspiration from the comprehensive frameworks of Hofstede, GLOBE, and others. Frameworks like these have been widely applied and, generally, for good reason. By aiming to capture key cultural differences across groups and societies in a limited number of dimensions they help us think about how exactly preferences differ and how they relate to situations people face in their everyday lives. This aspect of culture frameworks is not in dispute—it is just that the empirical evidence on the validity of SP data leads me to find these frameworks wanting in meaningful empirical content. My second main suggestion for future research therefore is to apply the EP method in the construction of comprehensive quantitative frameworks of cultural differences between societies and other groups along the lines of Hofstede (1980, 2001) or GLOBE, but based on EP measures of preference heterogeneity.

To see how this would work it is useful to describe briefly how Hofstede (1980), who provided the blueprint for many projects to follow, went about constructing his indices of cultural differences. Hofstede's indices are based on answers given in IBM's international employee attitude survey program. Between 1967 and 1973, two survey rounds rendered approximately 116,000 questionnaires from 72 countries (Hofstede 2001). For his indices, Hofstede aggregated individual responses at the country level and then applied data-reduction techniques, viz. factor analysis. Dropping countries with less than 50 respondents and questionnaire items deemed of poor quality, he was able to identify four—later five—dimensions of cultural values (Hofstede 2001: xix-xx).

An EP-based framework seeking to capture key cultural differences between societies could follow the same basic approach. A most practical start is to use a dataset like the World Values Survey or the European Social Survey.[68] Drawing on the large literature on the determinants of SWB, experienced preferences may be measured for a range of factors,

[68] The Gallup World Poll (see Chapter 4) would cover more societies but it is not publicly available and therefore does not allow for independent replication. I used data from the World Values Survey and the European Social Survey in my empirical assessment of the psychometric quality of EP measures of preference heterogeneity in Chapter 5.

employment status, income, leisure, occupational status, et cetera. These can be linked to likely dimensions of culture, such as the ones identified by Hofstede (1980, 2001), beforehand, but his is not a prerequisite. Different empirical methods can be used, ranging from estimating country-specific experienced utility functions or applying a multilevel technique with random coefficients and obtain posterior estimates for the slopes—I applied both strategies in Chapter 5 without finding much difference. Along the way, a large amount of judgment is involved: which determinants does one include? Is there a preferred SWB indicator, and how should one specify the models to be estimated? Et cetera. I do not think there is only a single correct recipe. Because rigor or comprehensiveness in one area is bound to come at the expense of rigor or comprehensiveness in another area, a priori, I do not want to argue that one particular strategy, technique, or model is better than another. What is more, all these questions concern only the first stage of estimating heterogeneous experienced utility functions to obtain country-level EP measures. I envision a second stage that brings together the data assembled in the first stage and turns them into a reliable and valid framework measuring key differences in cultural values between societies or other groups with a shared background. Data reduction techniques can be applied to establish patterns that are not immediately apparent, but still a large amount of judgment is involved. I can only speculate on how many dimensions may be uncovered or on which particular analytical technique will be most efficient in extracting as much information as possible from the data and in organizing them in the most accessible and intuitive manner. Ultimately, the proof of the pudding is in the eating.

As indicated in the previous section, the potential for applying the EP method in the construction of comprehensive quantitative frameworks of cultural differences between groups may be limited by the method's requirement of a sufficient amount of variation in the determinants of SWB within each group. I think a solution may be found in refraining, in first instance, from identifying groups beforehand and applying a two-step procedure instead; this would be before the results are analyzed further. The first step would involve a latent class technique to endogenously establish heterogeneity in the relation between SWB and one of its determinants, e.g. income, occupational status, et cetera. Clark et al. (2005) do this for the relation between income and financial satisfaction. They uncover four classes (with class membership based on the way income affects satisfaction with one's finances) but obviously these numbers may differ for the various determinants of SWB: for income we may uncover five classes and for occupational status only three. The second step would be to relate the complete set of classes thus identified to the groups of interest, most likely countries. Clark et

al. (2005) find that nationality is a strong predictor of class membership, and we could assess the likelihood of a citizen of a particular country belonging to a particular class for the complete set of classes indentified in the first step. This way we obtain a large collection of probabilities for all countries in the sample, and these would be the empirical input to the second stage in which, as mentioned above, data reduction techniques can be applied to establish systematic patterns in the country differences in preferences.

Whether the need for variation limits the EP approach is ultimately a practical matter. Similarly, the value of applying a latent class technique needs to be proven in practice. Nevertheless, I look forward to wider applications of the EP method, not least in the construction of comprehensive frameworks of cultural differences in preferences (as just described).

8.4.2. The Evolutionary and Biological Basis of Heterogeneous Preferences

The Evolutionary Foundations of (Heterogeneous) Preferences

In their de gustibus paper Stigler and Becker (1977: 76) stated, tongue-in-cheek, that preferences or tastes might be studied by the likes of psychologists, anthropologists, phrenologists, or sociobiologists (but not by economists). Their jibe may not have been far off the mark. More and more research appears devoted to a better understanding of human nature and variation therein. The earliest proponent most likely was Hippocrates (ca. 400 BC), whose *On Air, Water, and Places* provides an extensive study of climate, including its influence on the "constitutional tendencies" of various European and Asian populations.

One of the developing areas seeking to illuminate and explain preferences and other mental traits in humans is evolutionary psychology (and/or the closely related disciplines of human behavioral ecology and sociobiology) (e.g. Tooby and Cosmides 1992; see, also, Daly and Wilson 2001; Hames 2001; Wilson 1975; Winterhalder and Smith 2000). Key question in this field concerns the evolutionary foundations of a variety of tastes and behaviors, for instance male and female preferences for physical attractiveness of a mate (Buss and Barnes 1986; Buss 1989) or the general preference for Savanna-like natural landscapes (Falk and Balling 2010; Kaplan 1987; Orians and Heerwagen 1992) (see Buss 1995 and Cronk 1991 for reviews; Robson 2001, 2002 argues for consideration of such work in economics). The thrust of this type of work is that one can formulate an evolutionary process that has led humans to prefer certain characteristics in a mate or certain features of their habitat. Thus, a Savanna

landscape offers several evolutionary advantages; notably it has a lot of open space which facilitates the detection of potential threats, whilst the bushes and low trees provide refuge.

The study of the evolutionary foundations of preferences does not stop with uncovering the fitness logic of a particular behavior or the associated psychological mechanism, however. More relevant to the theme of this thesis, extensions have looked at how variation in preferences may emerge through evolutionary selection and turn into stable cross-cultural differences. The idea is that people's (natural) environment affects their preferences through the same kind of evolutionary pressures that shape preferences more generally. This way these extensions give a detailed analysis of how and why preferences differ between well-defined classes or groups, specifically between individuals from different natural environments. Particularly prominent in this regard is the work that seeks to understand cross-cultural differences as a resultant of differences in infectious disease and pathogen prevalence (see, for instance, Fincher et al. 2008 and Schaller and Murray 2008; Gangestad et al. 2006 and Schaller 2006 provide overviews).

Within the focus of this thesis, the most important thing to note is that this type of work may find useful application in the empirical study of the economic consequences of heterogeneous preferences, notably as an instrument for preference heterogeneity. Below I expand on this possibility previously alluded to in the context of the endogeneity of preferences (see above). First, however, I briefly discuss work assessing the biological basis of preference heterogeneity, which complements evolutionary studies of heterogeneous preferences.

The Biology of Preference Heterogeneity

Partly in parallel with work in evolutionary psychology, the last couple of years increasing attention has been paid to the biological (or neuroanatomical or physiological) basis of preference heterogeneity. Within the developing biology of heterogeneous preferences literature different strands can be discerned. Some work focuses on genetic issues such as heritability of preferences and the link between an individual's genetic make-up and his or her behavior, but there is also work examining the neurochemistry of risk preferences and trust behavior for example.

The first (and most) explicit linking of biology and preference heterogeneity can be found in studies of the neurophysiologic and genetic foundations of personality traits (measured through questionnaires). This work typically finds that about 50% of the variation in personality traits is heritable and the remaining part due to environment (Bouchard and

McGue 2003, McGue and Bouchard 1998; also see Loehlin 1992 and Plomin et al. 1994). In terms of specific genes, much research has looked at the serotonin transporter gene 5-HTTLPR (also known as the "Prozac" gene), relating it to personality traits (e.g. Lesch et al. 1996; Schinka et al. 2004; Sen et al. 2004) or cultural differences, notably in individualism-collectivism (Chiao and Blizinsky 2010; Way and Liebermann 2010). Chiao's (2010) foreword to the special issue on "cultural neuroscience" in *Social Cognitive & Affective Neuroscience* provides a further introduction to this latter type of research.

Evolutionary and Biological Instruments for Preference Heterogeneity

The discussion above has highlighted, amongst others, that variation in preferences tends to have an evolutionary foundation in differences in past environmental circumstances, notably historical pathogen incidence, and/or a biological foundation in neurophysiological differences between individuals and collectives. These relationships suggest that data on either of these foundations can supply researchers with a proxy for cultural differences in preferences. Researchers may not need to measure preference heterogeneity but could use these evolutionary or biological proxies instead. The biggest advantage of doing so would be that these proxies provide an exogenous instrument for cultural differences in preferences, and thus, as indicated above, may serve to overcome the challenge posed by the endogeneity of preferences, particularly in studies seeking a role for preference heterogeneity in economic differences.

As far as I know, the recent study by Gorodnichenko and Roland (2010) is the only one in economics taking this approach. They analyze the effect of individualism/collectivism on economic performance, using Chiao and Blizinsky's (2010) and Way and Liebermann's (2010) genetic data and Fincher et al.'s (2008) data on historical pathogen prevalence as instruments.[69] Because markets and other institutions do not affect the historical incidence of infectious disease or the frequency of particular gene types in a given population, empirical analyses using these indicators are less prone to suffer the limitations posed by such endogeneity of preferences. Therefore, compared to studies using RP, SP and/or EP measures, this work may be better able to shed light on the role of preference heterogeneity in differences in socioeconomic outcomes and developments.

[69] An interesting example in biology is Letendre et al.'s study (2010) of the intensity of infectious disease as a factor explaining cross-country variation in the frequency of intrastate armed conflict and civil war.

We should note, however, that substituting evolutionary or biological proxies for quantitative measures of preference heterogeneity is not without its own set of limitations. Firstly, environmental conditions such as pathogen stress likely have a strong direct effect on economic performance, next to a possible indirect effect that runs through evolved cultural traits (cf. Easterly and Levine 2003). If this is the case, the end result is that the endogeneity problem has been replaced by another challenging problem, namely the separation of the direct and indirect effect of environment on social and economic outcomes.[70] Secondly, and more importantly, the significance of a measure of, say, historical pathogen prevalence, does not yet specify what exactly it is about societies' cultures that feeds economic disparities. A necessary condition is that it should be clear for which particular cultural trait the pathogen-based measure (or another evolutionary instrument) is supposed to act as a proxy, individualism, collectivism, et cetera? This would require, at the least, a clear evolutionary rationale as to why a particular condition has fostered a certain trait. Even better would be to have an extensive set of reliable and valid measures of preference heterogeneity by which to establish the cultural traits proxied for by measures of pathogen prevalence or other evolutionary instruments. Failure to achieve this would severely hamper any contribution the use of evolutionary proxies for preference heterogeneity may make to our understanding of the economics of preference heterogeneity. Overall, direct measures of preference heterogeneity are preferred because they have a greater potential of illuminating the exact role of preference heterogeneity in economic differences, whilst evolutionary proxies are most insightful in cases in which the direct measures are so contaminated by endogeneity issues that they no longer seem valid.

[70] To make things even more complicated, variation in both pathogen richness and pathogen prevalence is partly accounted for by such factors as the number of birds and mammal species in a region (richness) and disease control efforts (Dunn et al. 2010). Other work finds that the incidence of infectious disease explains national differences in average IQ scores, and that it remains the most powerful explanatory variable after other factors such as temperature, gross domestic product per capita and level of education are controlled for (Eppig et al. 2010). At the least such studies suggest that it is not obvious what exactly correlations between measures of pathogen prevalence (either historically or contemporary) and socioeconomic outcomes signify.

8.5. CONCLUSION

Over the last two decades or so a distinct literature on the economics of preference heterogeneity has been emerging. A telling illustration is the foundation in 1997 of The Norms and Preferences Network, for which a group of economists including George Akerlof, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Edward Glaeser, Avner Greif, David Laibson, George Loewenstein, Casey Mulligan and Paul Romer teamed up with anthropologists and other social scientists with the aim of bringing together people with different backgrounds for the interdisciplinary study of preferences. This thesis fits this network and its goal, further resonating with such fields as new institutional economics, behavioral economics and (cross-cultural) psychology. It has focused on the different methods by which differences in preferences between groups may be given empirical content. Its main contribution concerns the development of a particular approach to measuring preferences and assessing preference heterogeneity, which involves estimating experienced utility functions the shape and content of which is allowed to differ between classes. The thesis demonstrated the value of this *experienced preference* approach, not least by using it to shed new light on two long-standing issues in the economics of preference heterogeneity, namely the extent to which Protestantism is associated with a specific work ethic and whether an economically debilitating communist value legacy may account for persistent differences in economic performance between the East and the West, specifically in Germany. Overall the findings in the thesis strongly put forward the experienced preference method as a most insightful approach to studying the economics of preference heterogeneity.

References

A

Adler, M., and E.A. Posner. 2008. Happiness Research and Cost-Benefit Analysis. *Journal of Legal Studies*, 37 (2): S253-S292.

Ailon, G. 2008. Mirror, mirror on the wall: Culture's consequences in a value test of its own design. *Academy of Management Review*, 33 (4): 885-904.

Akerlof, G.A. 2002. Behavioral Macroeconomics and Macroeconomic Behavior. *American Economic Review*, 92 (3): 411-433.

Akerlof, G.A. 2007. The Missing Motivation in Macroeconomics. *American Economic Review*, 97 (1): 5-36

Akerlof, G.A., and R.E. Kranton. 2000. Economics and Identity. *Quarterly Journal of Economics*, 115 (3): 715-753.

Akerlof, G.A., and R.E. Kranton. 2002. Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40 (4): 1167-1201.

Akerlof, G.A., and R.E. Kranton. 2005. Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19 (1): 9-32.

Akerlof, G.A., and R.E. Kranton. 2008. Identity, Supervision, and Work Groups. *American Economic Review, Papers and Proceedings*, 98 (2): 212-217.

Akerlof, G.A., and R.E. Kranton. 2009. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton, NJ: Princeton University Press.

Akerlof, G.A., and R.J. Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton, NJ: Princeton University Press.

Aknin, L.B., C.P. Barrington-Leigh, E.W. Dunn, J.F. Helliwell, R. Biswas-Diener, I. Kemeza, P. Nyende, C.E. Ashton-James, and M.I. Norton. 2010. *Prosocial Spending and Well-Being: Cross-Cultural Evidence for a Psychological Universal*. NBER Working Paper No. 16415. Cambridge, MA: National Bureau of Economic Research.

Alatas, S.H. 1977. *The myth of the lazy native*. London: Frank Cass.

Alesina, A., R. Di Tella, and R. MacCulloch. 2004. Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics*, 88 (9-10): 2009-2042.

Alesina, A., and N. Fuchs-Schündeln. 2007. Good Bye Lenin (or Not?)—The Effect of Communism on People's Preferences. *American Economic Review*, 97 (4): 1507-1528.

Alesina, A., and E. La Ferrara. 2005. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89 (5-6): 897-931.

Allport, G.W., and P.E. Vernon. 1931. *Study of values: Manual*. Boston, MA: Houghton Mifflin.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman. 2005. How much do we care about absolute versus relative income and consumption? *Journal of Economic Behavior & Organization*, 56 (3): 405-421.

Álvarez-Díaz, Á., L. González, and B. Radcliff. 2010. The Politics of Happiness: On the Political Determinants of Quality of Life in the American States. *Journal of Politics*, 72 (3): 894-905.

American Educational Research Association, the American Psychological Association, and the U.S. National Council on Measurement in Education. 1999. *Standards for educational and psychological testing*. Washington, DC: Author.

Anckar, D. 2002. Why are small island states democracies? *The Round Table*, 365 (1): 375-390.

Andersen, S., E. Bulte, U. Gneezy, and J.A. List. 2008. Do Women Supply More Public Goods than Men? Preliminary Experimental Evidence from Matrilineal and Patriarchal Societies. *American Economic Review, Papers and Proceedings*, 98 (2): 376-381.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström. 2008a. Eliciting Risk and Time Preferences. *Econometrica*, 76 (3): 583-618.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström. 2008b. Lost in state space: Are preferences stable? *International Economic Review*, 49 (3): 1091-1112.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström. 2010. Preference Heterogeneity in Experiments: Comparing the Field and Laboratory. *Journal of Economic Behavior & Organization*, 73 (2): 209-224.

Anderson, L.R., and J.M. Mellor. 2009. Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, 39 (2): 137-160.

Andreoni, J., and L. Vesterlund. 2001. Which is the Fair Sex? Gender Differences in Altruism. *Quarterly Journal of Economics*, 116 (1): 293-312.

Andrews, F. 1984. Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *Public Opinion Quarterly*, 48 (2): 409-442.

Andrews, F.M., and R. Crandall. 1976. The validity of measures of self-reported well-being. *Social Indicators Research*, 3 (1): 1-19.

Andrews, F.M., and S.B. Withey. 1976. *Social indicators of well-being*. New York and London: Plenum Press.

Angner, E. 2005. *Subjective measures of well-being: A philosophical examination*. PhD thesis, Pittsburgh: University of Pittsburgh.

Antecol, H. 2000. An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates. *Labour Economics*, 7 (4): 409-426.

Antecol, H. 2001. Why Is There Interactive Variation in the Gender Wage Gap? The Role of Cultural Factors. *Journal of Human Resources*, 36 (1): 119-143.

Argyle, M. 1999. Causes and Correlates of Happiness. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation: 353-373.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. *Report of the NOAA panel on contingent valuation*. Electronic version available from www.darp.noaa.gov/library/pdf/cvblue.pdf.

Arruñada, B. 2010. Protestants and Catholics: Similar Work Ethic, Different Social Ethic. *Economic Journal*, 120 (547): 890-918.

Atkinson, S.M., S.B. Baird, and M.B. Frye. 2003. Do female fund managers manage differently. *Journal of Financial Research*, 26 (1): 1-18.

B

Bandura, A. 1999. Social cognitive theory of personality. In: Pervin, L.A., and O.P. John, Eds. *Handbook of personality: Theory and research*. New York: Guilford.

Barber, B., and T. Odean. 2001. Boys Will Be Boys: Gender, Overconfidence and Common Stock Investment. *Quarterly Journal of Economics*, 116 (1): 261-292.

Barro, R.J. 1996. Democracy and growth. *Journal of Economic Growth*, 1 (1): 1-27.

Barro, R.J. 1999. Determinants of Democracy. *Journal of Political Economy*, 107 (6): S158-S183.

Barro, R.J., and R.M. McCleary. 2003. Religion and economic growth across countries. *American Sociological Review*, 68 (5): 760-781.

Baskerville, R.F. 2003. Hofstede never studied culture. *Accounting, Organizations and Society*, 28 (1): 1-14.

Baskerville, R.F. 2005. A research note: the unfinished business of culture. *Accounting, Organizations and Society*, 30 (4): 389-391.

Becker, G.S. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Becker, G.S. 1996. *Accounting for Tastes*. Cambridge, MA: Harvard University Press.

Becker, S.O., and L. Woessmann. 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics*, 124 (2): 531-596.

Becker, S.O., and L. Woessmann. 2010. The Effect of Protestantism on Education before the Industrialization: Evidence from 1816 Prussia. *Economics Letters*, 107 (2): 224-228.

Beckmann, D., L. Menkhoff, and M. Suto. 2008. Does culture influence asset managers' views and behavior? *Journal of Economic Behavior & Organization*, 67 (3-4): 624-643.

Bellemare, C., S. Kröger, and A. van Soest. 2008. Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica*, 76 (4): 815-839.

Benjamin, D.J., C.F. Chabris, E.L. Glaeser, V. Gudnason, T.B. Harris, D.I. Laibson, L. Launer, and S. Purcell. 2008. Genoeconomics. In: Weinstein, M., J.W. Vaupel, and K.W. Wachter, Eds. *Biosocial Surveys*. Committee on Population, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Berg, J., J. Dickhaut, and K. McCabe. 2005. Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences*, 102 (11): 4209-4214.

Bernheim, B.D. 2009. On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal. *American Economic Journal: Microeconomics*, 1 (2): 1-41.

Bertrand, M., and S. Mullainathan. 2001. Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review, Papers and Proceedings*, 91 (2): 67-72.

Beshears, J., J.J. Choi, D. Laibson, and B.C. Madrian. 2008. How are preferences revealed? *Journal of Public Economics*, 92 (8-9): 1787-1794.

Beugelsdijk, S., and B. Frijns. 2010. A cultural explanation of the foreign bias in international asset allocation. *Journal of Banking and Finance*, 34 (9): 2121-2131.

Beugelsdijk, S., and R. Maseland. 2011. *Culture in Economics. History, Methodological Reflections, and Contemporary Applications*. Cambridge: Cambridge University Press.

Beugelsdijk, S., and R. Smeets. 2008. Entrepreneurial culture and economic growth; revisiting McClelland's thesis. *American Journal of Economics and Sociology*, 67 (5): 915-939.

Bjørnskov, C., N. Datta Gupta, and P.J. Pedersen. 2008. Analyzing trends in subjective well-being in 15 European countries, 1973-2002. *Journal of Happiness Studies*, 9 (2): 317-330.

Bjørnskov, C., A. Dreher, and J.A.V. Fischer. 2008. On Decentralization and Life Satisfaction. *Economics Letters*, 99 (1): 147-151.

Bjørnskov, C., A. Dreher, and J.A.V. Fischer. 2010. Formal institutions and subjective well-being: Revisiting the cross-country evidence. *European Journal of Political Economy*, 26 (4): 419-430.

Blackburn, S. 1996. happiness. In: *The Oxford Dictionary of Philosophy*. Oxford University Press: Oxford Reference Online.

Blanchflower, D.G., and R. Freeman. 1997. The attitudinal legacy of Communist labor relations. *Industrial and Labor Relations Review*, 50 (3): 438-459.

Blanchflower, D.G., and A.J. Oswald. 2008. Hypertension and happiness across nations. *Journal of Health Economics*, 27 (2): 218-233.

Blau, F.D., L.M. Kahn, and K.L. Papps. 2011. Gender, Source Country Characteristics and Labor Market Assimilation Among Immigrants: 1980-2000. *Review of Economics and Statistics*, doi:10.1162/REST_a_00064.

Bok, D. 2010. *The Politics of Happiness: What Government Can Learn from the New Research on Well-Being*. Princeton, NJ: Princeton University Press.

Boltho, A., W. Carlin, and P. Scaramozzino. 1997. Will East Germany become a new Mezzogiorno? *Journal of Comparative Economics*, 24 (3): 241-264.

Bolton, G.E., and E. Katok. 1995. An experimental test for gender differences in beneficent behavior. *Economics Letters*, 48 (3-4): 287-292.

Bond, M.H. 1988. Invitation to a wedding: Chinese values and economic growth. In: Sinha, D., and H.S.R. Kao, Eds. *Social Values and Development: Asian Perspectives*. New Delhi: Sage Publications.

Bond, M.H. 2002. Reclaiming the Individual From Hofstede's Ecological Analysis—A 20-Year Odyssey: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128 (1): 73-77.

Borghans, L., A.L. Duckworth, J.J. Heckman, and B. ter Weel. 2008. The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43 (4): 972-1059.

Borghans, L., B.H.H. Golsteyn, J.J. Heckman, and H. Meijers. 2009. Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association*, 7 (2-3): 649-658.

Boskin, M.J., and E. Sheshinski. 1978. Optimal Redistributive Taxation When Individual Welfare Depends Upon Relative Income. *Quarterly Journal of Economics*, 92 (4): 589-600.

Bouchard, T.J., Jr., and M. McGue. 2003. Genetic and Environmental Influences on Human Psychological Differences. *Journal of Neurobiology*, 54 (1): 4-45.

Boulding, K.E. 1966. The Economics of Knowledge and the Knowledge of Economics. *American Economic Review*, 56 (1/2): 1-13.

Boulding, K.E. 1972. Human Betterment and the Quality of Life. In: Burkhard Strumpel, J.N. Morgan, and E. Zahn, Eds. *Human Behavior in Economic Affairs: Essays in honor of George Katona*. Amsterdam: Elsevier.

Bowles, S. 1998. Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature*, 36 (1): 75-111.

Boyd, R. 2008. cross-cultural experiments. In: Durlauf, S.N., and L.E. Blume, Eds. *The New Palgrave Dictionary of Economics*. Second Edition. Online. Palgrave Macmillan.

Brandstätter, H. 1997. Becoming an entrepreneur – A question of personality structure? *Journal of Economic Psychology*, 18 (2-3): 157-177.

Bratton, M. 2003. Briefing: Islam, democracy, and public opinion in Africa. *African Affairs*, 102 (408): 493-501.

Brewer, P., and S. Venaik. 2010. GLOBE practices and values – A case of diminishing marginal utility? *Journal of International Business Studies*, 41 (8): 1316-1324.

Brockhaus, R.H. 1982. The psychology of an entrepreneur. In: Kent, C., D.L. Sexton, and K.H. Vesper, Eds. *Encyclopedia of entrepreneurship*. Englewood Cliffs, NJ: Prentice Hall.

Brodbeck, F.C., and M. Frese. 2007. Societal culture and leadership in Germany. In: Chhokar, J.S., F.C. Brodbeck, and R.J. House, Eds. *Culture and leadership across the world: The GLOBE book of in-depth studies*. Mahwah, NJ: Lawrence Erlbaum.

Burlando, R., and J.D. Hey. 1997. Do Anglo-Saxons free ride more? *Journal of Public Economics*, 64 (1): 41-60.

Burnham, T.C., and R. Kurzban. 2005. On the limitations of quasi-experiments. Comment on Henrich et al.: Economic behavior in cross-cultural perspective. *Behavioral and Brain Sciences*, 28 (6): 818-819.

Buss, D.M. 1989. Sex differences in human mate preferences: Evolutionary hypotheses testing in 37 cultures. *Behavioral and Brain Sciences*, 12 (1): 1-14.

Buss, D.M. 1995. Evolutionary Psychology: A New Paradigm for Psychological Science, *Psychological Inquiry*, 6 (1): 1-30.

Buss, D.M., and M. Barnes. 1986. Preferences in Human Mate Selection. *Journal of Personality and Social Psychology*, 50 (3): 559-570

C

Camerer, C. 1999. Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96 (19): 10575-10577.

Camerer, C., and E. Fehr. 2004. Measuring social norms and preferences using experimental games: A guide for social scientists. In: Henrich, J., R. Boyd, et al., Eds. *Foundations of Human Sociality*. Oxford: Oxford University Press.

Camerer, C.F, and G. Loewenstein. 2003. Behavioral Economics: Past, Present, Future. In: Camerer, C.F, G. Loewenstein, and M. Rabin, Eds. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.

Camerer, C.F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Camerer, C.F. 2007. Neuroeconomics: Using Neuroscience to Make Economic Predictions. *Economic Journal*, 117 (519): C26-C42.

Camerer, C.F., G. Loewenstein, and D. Prelec. 2004. Neuroeconomics: Why Economics Needs Brains. *Scandinavian Journal of Economics*, 106 (3): 555-579.

Camerer, C.F., G. Loewenstein, and D. Prelec. 2005. Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature*, 43 (1): 9-64.

Cameron, T.A. 2008. contingent valuation. In: Durlauf, S.N., and L.E. Blume, Eds. *The New Palgrave Dictionary of Economics*. Second Edition. Online. Palgrave Macmillan.

Campbell, D.T. 1957. Factors Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin*, 54 (4): 297-312.

Campbell, D.T., and J.C. Stanley. 1966. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally.

Caplan, B. 2003. Stigler-Becker versus Myers-Briggs: Why preference-based explanations are scientifically meaningful and empirically important. *Journal of Economic Behavior & Organization*, 50 (4): 391-405.

Caprara, G.V., and D. Cervone. 2000. *Personality: Determinants, Dynamics, and Potentials*. Cambridge: Cambridge University Press.

Carroll, N., P. Frijters, and M. Shields. 2009. Quantifying the costs of drought: new evidence from life satisfaction data. *Journal of Population Economics*, 22 (2): 445-461.

Central Intelligence Agency. 2009. *The World Factbook 2009*. Washington, DC: Central Intelligence Agency.

Chang, H.-J. 2007. *Bad Samaritans – Rich Nations, Poor Policies, and the Threat to the Developing World*. London: Random House.

Chen, M.K. 2008. *Rationalization and cognitive dissonance. Do choices affect or reflect preferences?* Cowles Foundation Discussion Paper No. 1669. Cowles Foundation for Research in Economics, Yale University.

Chen, M. K., and J.L. Risen. 2009. Is choice a reliable predictor of choice? A comment on Sagarin and Skowronski. *Journal of Experimental Social Psychology*, 45 (2): 425-427.

Chhokar, J.S., F.C. Brodbeck, and R.J. House, Eds. 2007. *Culture and leadership across the world: The GLOBE book of in-depth studies*. Mahwah, NJ: Lawrence Erlbaum.

Chiao, J.Y. 2010. Editorial. At the frontier of cultural neuroscience: Introduction to the special issue. *Social Cognitive & Affective Neuroscience*, 5 (2-3): 109-110.

Chiao, J.Y., and K.D. Blizinsky. 2010. Culture-gene coevolution of individualism-collectivism and the serotonin transporter gene. *Proceedings of the Royal Society B: Biological sciences*, 277 (1681): 529-537.

Chuah, S.-H., R. Hoffmann, M. Jones, and G. Williams. 2007. Do cultures clash? Evidence from cross-national ultimatum game experiments. *Journal of Economic Behavior & Organization*, 64 (1): 35-48.

Chuah, S.-H., R. Hoffmann, M. Jones, and G. Williams. 2009. An economic anatomy of culture: Attitudes and behaviour in inter- and intra-national ultimatum game experiments. *Journal of Economic Psychology*, 30 (5): 732-744.

Chui, A.C.W., S. Titman, and J. Wei. 2010. Individualism and Momentum around the World. *Journal of Finance*, 65 (1): 361-392.

Citro, C.F., and R.T. Michael, Eds. 1995. *Measuring Poverty: A New Approach*. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, National Research Council, Washington, DC: National Academy Press.

Clague, C., S. Gleason, and S. Knack. 2001. Determinants of lasting democracy in poor countries: Culture, development, and institutions. *Annals of the American Academy of Political and Social Science*, 573 (1): 16-41.

Clark, A.E. 2003. Unemployment as a social norm: Psychological evidence from panel data. *Journal of Labor Economics*, 21 (2): 323-351.

Clark, A.E., E. Diener, Y. Georgellis, and R.E. Lucas. 2008. Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis. *Economic Journal*, 118 (529): F222-F243.

Clark, A.E., F. Etilé, F. Postel-Vinay, C. Senik, and K. Van der Straeten. 2005. Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries. *Economic Journal*, 115 (502): C118-C132.

Clark, A.E., P. Frijters, and M.A. Shields. 2008. Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, 46 (1): 95-144.

Clark, A.E., and O. Lelkes. 2005. *Deliver Us From Evil: Religion as Insurance*. Unpublished manuscript, Paris School of Economics.

Clark, A.E., and A.J. Oswald. 1994. Unhappiness and Unemployment. *Economic Journal*, 104 (424): 648-659.

Clark, A.E., and A.J. Oswald. 1996. Satisfaction and comparison income. *Journal of Public Economics*, 61 (3): 359-381.

Clark, A.E., and A.J. Oswald. 2002. A Simple Statistical Model for Measuring How Life Events Affect Happiness. *International Journal of Epidemiology*, 31 (6): 1139-1144.

Clark, G. 2007. *A Farewell to Alms: A Brief Economic History of the World*. Princeton, NJ: Princeton University Press.

Clarke, H.D., and N. Dutt. 1991. Measuring Value Change in Western Industrialized Societies: The Impact of Unemployment. *American Political Science Review*, 85 (3): 905-20.

Clarke, H.D., N. Dutt, and A. Kornberg. 1993. The Political Economy of Attitudes toward Polity and Society in Western European Democracies. *Journal of Politics*, 55 (4): 998-1021.

Clarke, H.D., N. Dutt, and J. Rapkin. 1997. Conversations in Context: The (Mis)Measurement of Value Change in Advanced Industrial Societies. *Political Behavior*, 19 (1): 19-40.

Clarke, H.D., A. Kornberg, C. McIntyre, P. Bauer-Kaase, and M. Kaase. 1999. The Effect of Economic Priorities on the Measurement of Value Change. *American Political Science Review*, 93 (3): 637-647.

Corbett, D. 2004. *Breaking Down the Wall in the Head*. Deutsche Welle, October 3, 2004.

Corneo, G., and H.P. Grüner. 2002. Individual preferences for political redistribution. *Journal of Public Economics*, 83 (1): 83-107.

Costa, P.T., Jr., and R.R. McCrae. 1988. Personality in Adulthood: A Six-Year Longitudinal Study of Self-Reports and Spouse Ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, 54 (5): 853-863.

Cowen, T. 1989. Are all tastes constant and identical? A critique of Stigler and Becker. *Journal of Economic Behavior & Organization*, 11 (1): 127-135.

Cramer, J.S., J. Hartog, N. Jonker, and C.M. van Praag. 2002. Low risk aversion encourages the choice for entrepreneurship: The empirical test of a truism. *Journal of Economic Behavior & Organization*, 48 (1): 29-36.

Cromie, S. 2000. Assessing entrepreneurial inclinations: some approaches and empirical evidence. *European Journal of Work and Organizational Psychology*, 9 (1): 7-30.

Cronbach, L.J., and G.C. Gleser. 1965. *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

Cronbach, L.J., and P.E. Meehl. 1955. Construct Validity in Psychological Tests. *Psychological Bulletin*, 52 (4): 281-302.

Cronk, L. 1991. Human Behavioral Ecology. *Annual Review of Anthropology*, 20: 25-53.

Croson, R., and N. Buchan. 1999. Gender and Culture: International Experimental Evidence from Trust Games. *American Economic Review, Papers and Proceedings*, 89 (2): 386-391.

Croson, R., and S. Gächter. 2010. The science of experimental economics. *Journal of Economic Behavior & Organization*, 73 (1): 122-131.

Croson, R., and U. Gneezy. 2009. Gender Differences in Preferences. *Journal of Economic Literature*, 47 (2): 448-474.

Crossley, T.F., and S. Kennedy. 2002. The reliability of self-assessed health status. *Journal of Health Economics*, 21 (4): 643-658.

D

Daly, M., and M. Wilson 2001. Sociobiology: Overview. In: Smelser, N.J., and P.B. Baltes, Eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier.

Darity, W., and A.H. Goldsmith. 1996. Social Psychology, Unemployment and Macroeconomics. *Journal of Economic Perspectives*, 10 (1): 121-140.

Davis, J.A. 1996. Review Essay - Paul R. Abramson and Ronald Inglehart. Value Change in Global Perspective. *Public Opinion Quarterly*, 60 (2): 322-331.

Davis, D.W., and C. Davenport. 1999. Assessing the Validity of the Postmaterialism Index. *American Political Science Review*, 93 (3): 649-364.

Davis, D.W., K.M. Dowley, and B.D. Silver. 1999. Postmaterialism in World Societies: Is It Really a Value Dimension? *American Journal of Political Science*, 43 (3): 935-362.

de Jong, E. 2009. *Culture and Economics: On Values, Economics and International Business*. London: Routledge.

de Jong, E., R. Smeets, and J. Smits. 2006. Culture and Openness. *Social Indicators Research*, 78 (1): 111-136.

Deaton, A. 2008. Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll. *Journal of Economic Perspectives*, 22 (1): 53-72.

Delacroix, J., and F. Nielsen. 2001. The beloved myth: Protestantism and the rise of industrial capitalism in nineteenth-century Europe. *Social Forces*, 80 (2): 509-553.

Dezutter, J., B. Soenens, and D. Hutsebaut. 2006. Religiosity and mental health: A further exploration of the relative importance of religious behaviors vs. religious attitudes. *Personality and Individual Differences*, 40 (4): 807-818.

Di Tella, R., R.J. MacCulloch, and A.J. Oswald. 2001. Preferences Over Inflation and Unemployment: Evidence From Surveys of Happiness. *American Economic Review*, 91 (1): 335-341.

Di Tella, R., R.J. MacCulloch, and A.J. Oswald. 2003. The Macroeconomics of Happiness. *Review of Economics and Statistics*, 95 (4): 809-828.

Di Tella, R., and R. MacCulloch. 2005. Partisan Social Happiness. *Review of Economic Studies*, 72 (2): 367-393.

Di Tella, R., and R. MacCulloch. 2006. Some Uses of Happiness Data in Economics. *Journal of Economic Perspectives*, 20 (1): 25-46.

Di Tella, R., J. Haisken-De New, and R. MacCulloch. 2010. Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior & Organization*, 76 (3): 834-52.

Diamond, P.A., and J.A. Hausman. 1994. Contingent Valuation – Is Some Number better than No Number? *Journal of Economic Perspectives*, 8 (4): 45-64.

Diamond, J.M., and J.A. Robinson. 2010. *Natural Experiments of History*. Harvard, MA: Harvard University Press.

Diener, E. 1984. Subjective Well-Being. *Psychological Bulletin*, 95 (3): 542-575.

Diener, E. 1994. Assessing Subjective Well-Being: Progress and Opportunities. *Social Indicators Research*, 31 (2): 103-157.

Diener, E. 2000. Subjective Well-Being. The Science of Happiness and a Proposal for a National Index. *American Psychologist*, 55 (1): 34-43.

Diener, E. 2006. Guidelines for national indicators of subjective well-being and ill-being. *Journal of Happiness Studies*, 7 (4): 397-404.

Diener, E., and M. Diener. 1995. Cross-Cultural Correlates of Life Satisfaction and Self-Esteem. *Journal of Personality and Social Psychology*, 68 (8): 653-663.

Diener, E., R.A. Emmons, R.J. Larsen, and S. Griffin. 1985. The satisfaction with life scale. *Journal of Personality Assessment*, 49 (1): 71-75.

Diener, E., C. Gohm, E. Suh, and S. Oishi. 2000. Similarity of the Relations between Marital Status and Subjective Well-Being Across Cultures. *Journal of Cross-Cultural Psychology*, 31 (4): 419-436.

Diener, E., D. Kahneman, W. Tov, and R. Arora. 2010. Income's Association with Judgments of Life Versus Feelings. In: Diener, E., D. Kahneman, and J.F. Helliwell, Eds. *International differences in well-being*. Oxford: Oxford University Press.

Diener, E., P. Kesebir, and R. Lucas. 2008. Benefits of Accounts of Well-Being—For Societies and for Psychological Science. *Applied Psychology: An International Review*, 57 (1): 37-53.

Diener, E., R. Lucas, U. Schimmack, and J.F. Helliwell. 2009. *Well-Being for Public Policy*. New York: Oxford University Press.

Diener, E., and S. Oishi. 2004. Are Scandinavians happier than Asians? Issues in comparing nations on subjective well-being. In: Columbus, F., Ed. *Asian Economic and Political Issues, Volume 10*. New York: Nova Science.

Diener, E., S. Oishi, and R.E. Lucas. 2003. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, 54: 403-425.

Diener, E., E. Sandvik, W. Pavot, and D. Gallagher. 1991. Response artifacts in the measurement of subjective well-being. *Social Indicators Research*, 24 (1): 35-56.

Diener, E., J.J. Sapyta, and E.M. Suh. 1998. Subjective Well-Being Is Essential to Well-Being. *Psychological Inquiry*, 9 (1): 33-37.

Diener, E., and M.E.P. Seligman. 2004. Beyond Money. Toward an Economy of Well-Being. *Psychological Science in the Public Interest*, 5 (1): 1-31.

Diener, E., and E.M. Suh. 1997. Measuring quality of life: Economic, social and subjective indicators. *Social Indicators Research*, 40 (1-2): 189-216.

Diener, E., and E.M. Suh 1999. National Differences in Subjective Well-Being. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Diener, E., and E.M. Suh. 2000. Measuring Subjective Well-Being to Compare the Quality of Life of Cultures. In: Diener, E., and E.M. Suh, Eds. *Culture and Subjective Well-Being*. Cambridge, MA and London: The MIT Press.

Diener, E., E.M. Suh, H. Smith, and L. Shao. 1995. National differences in reported subjective well-being: Why do they occur? *Social Indicators Research*, 34 (1): 7-32.

Diener, E., E.-M. Suh, R.E. Lucas, and H.L. Smith. 1999. Subjective Well-Being: Three Decades of Progress. *Psychological Bulletin*, 125 (2): 276-302.

Dolan, P., and D. Kahneman. 2008. Interpretations of Utility and Their Implications for the Valuation of Health. *Economic Journal*, 118 (525): 215-234.

Dolan, P., and M.P. White. 2007. How Can Measures of Subjective Well-Being Be Used to Inform Public Policy. *Perspectives on Psychological Science*, 2 (1): 71-85.

Dominitz, J., and A. van Soest. 2008. survey data, analysis of. In: Durlauf, S.N., and L.E. Blume, Eds. *The New Palgrave Dictionary of Economics*. Second Edition. Online. Palgrave Macmillan.

Duch, R.M., and M.A. Taylor. 1993. Postmaterialism and the Economic Condition. *American Journal of Political Science*, 37 (3): 747-779.

Dunn, R.D., T.J. Davies, N.C. Harris, and M.C. Gavin. 2010. Global Drivers of Human Pathogen Richness and Prevalence. *Proceedings of the Royal Society of London B: Biological Sciences*, 277 (1694): 2587-2595.

Dutt, A.K., and B. Radcliff, Eds. 2009. *Happiness, Economics and Politics: Towards a Multi-Disciplinary Approach*. Cheltenham: Edward Elgar.

Dynan, K.E., and E. Ravina. 2007. Increasing Income Inequality, External Habits, and Self-Reported Happiness. *American Economic Review, Papers and Proceedings*, 97 (2): 226-231.

E

Easterlin, R.A. 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In: David, P.A., and M.W. Reder, Eds. *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*. New York and London: Academic Press.

Easterlin, R.A. 2003. Explaining happiness. *Proceedings of the National Academy of Sciences*, 100 (19): 11177-11183.

Easterlin, R.A., and L. Angelescu. 2009. *Happiness and Growth the World Over: Time Series Evidence on the Happiness-Income Paradox*. IZA Discussion Paper No. 4060. Bonn: Institute for the Study of Labor.

Easterlin, R.A., L. Angelescu McVey, M. Switek, O. Sawangfa, and J. Smith Zweig. 2010. The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences*, 107 (52): 22463-22468.

Easterly, W., and R. Levine. 2003. Tropics, germs and crops: How endowments influence economic development. *Journal of Monetary Economics*, 50 (1): 3-39.

Eckel, C.C. 2008. gender differences (experimental evidence). In: Durlauf, S.N., and L.E. Blume, Eds. *The New Palgrave Dictionary of Economics*. Second Edition. Online. Palgrave Macmillan.

Eckel, C.C., and P.J. Grossman. 1996. The relative price of fairness: Gender differences in a punishment game. *Journal of Economic Behavior & Organization*, 30 (2): 134-158.

Eckel, C.C., and P.J. Grossman. 1998. Are Women Less Selfish Than Men?: Evidence from Dictator Games. *Economic Journal*, 108 (448): 726-735.

Eckel, C.C., and P.J. Grossman. 2001. Chivalry and Solidarity in Ultimatum Games. *Economic Inquiry*, 39 (2): 171-188.

Edgeworth, F.Y. 1881. *Mathematical Psychics. An essay on the application of mathematics to the moral sciences*. London: C. Kegan Paul and Co.

Ehrhardt, J.J., W.E. Saris, and R. Veenhoven. 2000. Stability of life-satisfaction over time: Analysis of change in ranks in a national population. *Journal of Happiness Studies*, 1 (2): 177-205.

Eisenberg, P., and P.F. Lazarsfeld. 1938. The psychological effects of unemployment. *Psychological Bulletin*, 35 (6): 358-390.

Ekman, P., R.J. Davidson, and W.V. Friesen. 1990. The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology*, 58 (2): 342-353.

Ekman, P., W.V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W.A. LeCompte, T. Pitcairn, P.E. Ricci-Bitti, K. Scherer, M. Tomita, and A. Tzavaras. 1987. Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. *Journal of Personality and Social Psychology*, 53 (4): 712-717.

Ellison, C.G. 1995. Race, Religious Involvement and Depressive Symptomatology in a Southeastern U.S. Community. *Social Science & Medicine*, 40 (11): 1561-1572.

Ellison, C.G., and D. Fan. 2008. Daily Spiritual Experiences and Psychological Well-being among US Adults. *Social Indicators Research*, 88 (2): 247-271.

Eppig, C., C.L. Fincher, and R. Thornhill. 2010. Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, doi:10.1098/rspb.2010.0973.

Epstein, G.S., and I.N. Gang. 2007. Understanding the development of fundamentalism. *Public Choice*, 132 (3-4): 257-271.

Esping-Andersen, G. 1989. The three political economies of the welfare state. *Canadian Journal of Sociology and Anthropology*, 2 (1): 10-36.

Esping-Andersen, G. 1990. *The three worlds of welfare capitalism*. Princeton, NJ: Princeton University Press.

Eungi Kim, A., and G.-S. Park. 2003. Nationalism, Confucianism, work ethic and industrialization in the South. *Journal of Contemporary Asia*, 33 (1): 37-49.

European Values Study Group and World Values Survey Association. 2006. *European and World Values Surveys four-wave integrated data file, 1981-2004 v.20060423*. Aggregate File Producer: ASEP/JDS, Madrid.

Evans, G., and P. Rose. 2007. Support for democracy in Malawi: Does schooling matter? *World Development*, 35 (5): 904-919.

F

Fagenson, E.A. 1993. Personal value systems of men and women entrepreneurs versus managers. *Journal of Business Venturing*, 8 (5): 409-430.

Falk, J.H., and J.D. Balling. 2010. Evolutionary Influence on Human Landscape Preference. *Environment and Behavior*, 42 (4): 479-493.

Federal Statistical Office Germany. 2010. *GENESIS-Online database*. https://www-genesis.destatis.de.

Fehr, E. 2009. Economics and biology of trust. *Journal of the European Economic Association*, 7 (2-3): 235-266.

Fellows, E.W. 1966. Happiness: A survey of research. *Journal of Humanistic Psychology*, 6 (1): 17-30.

Fernández, R. 2007. Women, Work, and Culture. *Journal of the European Economic Association*, 5 (2-3): 305-332.

Fernández, R. 2008. culture and economics. In: Durlauf, S.N., and L.E. Blume, Eds. *The New Palgrave Dictionary of Economics*. Second Edition. Online. Palgrave Macmillan.

Fernández, R. 2010. *Does Culture Matter?* NBER Working Paper No. 16277. Cambridge, MA: National Bureau of Economic Research.

Fernández, R., and A. Fogli. 2006. Fertility: The Role of Culture and Family Experience. *Journal of the European Economic Association*, 4 (2-3): 552-561.

Fernández, R., and A. Fogli. 2009. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics*, 1 (1): 146-177.

Ferrer-i-Carbonell, A., and P. Frijters. 2004. How Important Is Methodology for the Estimates of the Determinants of Happiness? *Economic Journal*, 114 (497): 641-659.

Fidrmuc, J.P., and M. Jacob. 2010. Culture, Agency Costs, and Dividends. *Journal of Comparative Economics*, 38 (3): 321-339.

Fincher, C.L., R. Thornhill, D.R. Murray, and M. Schaller. 2008. Pathogen prevalence predicts human cross-cultural variability in individualism / collectivism. *Proceedings of the Royal Society B: Biological Sciences*, 275 (1640): 1279-1285.

Finkelstein, A., E.F.P. Luttmer, and M.J. Notowidigdo. 2008. *What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption*. NBER Working Paper No. 14089. Cambridge, MA: National Bureau of Economic Research.

Finkelstein, A., E.F.P. Luttmer, and M.J. Notowidigdo. 2009. Approaches to Estimating the Health State Dependence of the Utility Function. *American Economic Review, Papers and Proceedings*, 99 (2): 116-121.

Fischhoff, B. 1991. Value Elicitation: Is There Anything in There? *American Psychologist*, 46 (8): 835-847.

Fish, M.S. 2002. Islam and authoritarianism. *World Politics*, 55 (1): 4-37.

Fisher, I. 1906. [1930]. *The Nature of Capital and Income*. London: Macmillan & Co.

Fisher, I. 1918. Is "Utility" the Most Suitable Term for the Concept it is Used to Denote? *American Economic Review*, 8 (2): 335-337.

Fiske, A.P. 2002. Using individualism and collectivism to compare cultures-A critique of the validity and measurement of the constructs: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128 (1): 78-88.

Fitzpatrick, R., A. Fletcher, S. Gore, D. Jones, D. Spiegelhalter, and D. Cox. 1992. Quality of life measures in health care. I: Applications and issues in assessment. *British Medical Journal*, 305 (6861): 1074-1077.

Fleurbaey, M. 2009. Beyond GDP: The Quest for a Measure of Social Welfare. *Journal of Economic Literature*, 47 (4): 1029-1075.

Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. Elger, and A. Falk. 2007. Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum. *Science*, 318 (23 November): 1305-1308.

Foster, J.E. 1998. Absolute versus Relative Poverty. *American Economic Review, Papers and Proceedings*, 88 (2): 335-341.

Frank, R.H. 1999. *Luxury Fever: Money and Happiness in an Era of Excess*. New York: The Free Press.

Frank, R.H. 2005. Positional Externalities Cause Large and Preventable Welfare Losses. *American Economic Review, Papers and Proceedings*, 95 (2): 137-141.

Franke, R.H., G. Hofstede, and M.H. Bond. 1991. Cultural Roots of Economic Performance: A Research Note. *Strategic Management Journal*, 12 (summer): 165-173.

Frankel, S.H. 1953. *The economic impact on under-developed societies. Essays on International Investment and Social Change*. Reprinted April 1955. Oxford: Basil Blackwell.

Frey, S.R. 1984. Does N-achievement cause economic development? A cross-lagged panel analysis of the McClelland thesis. *Journal of Social Psychology*, 122 (1): 67-70.

Frey, B.S., S. Luechinger, and A. Stutzer. 2007. Calculating tragedy: Assessing the costs of terrorism. *Journal of Economic Surveys*, 21 (1): 1-24.

Frey, B.S., S. Luechinger, and A. Stutzer. 2009. The life satisfaction approach to valuing public goods: The case of terrorism. *Public Choice*, 138 (3-4): 317-345.

Frey, B.S., S. Luechinger, and A. Stutzer. 2010. The Life Satisfaction Approach to Environmental Valuation. *Annual Review of Resource Economics*, 2: 139-160.

Frey, B.S., and A. Stutzer. 2000. Happiness, Economy and Institutions. *Economic Journal*, 110 (466): 918-938.

Frey, B.S., and A. Stutzer. 2002. What Can Economists Learn from Happiness Research? *Journal of Economic Literature*, 40 (2): 402-435.

Frey, B.S., and A. Stutzer. 2010. Happiness and public choice. *Public Choice*, 144 (3-4): 557-573.

Freytag, A., and Thurik, R. 2007. Entrepreneurship and its determinants in a cross country setting. *Journal of Evolutionary Economics*, 17 (2): 117-131.

Friedman, M. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Fryer, D. 2001. Unemployment and Mental Health. In: Smelser, N.J., and P.B. Baltes, Eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier.

Fuchs-Schündeln, N. 2008. The Response of Household Saving to the Large Shock of German Reunification. *American Economic Review*, 98 (5): 1798-1828.

Fuchs-Schündeln, N., and M. Schündeln. 2005. Precautionary Savings and Self-Selection: Evidence from the German Reunification "Experiment". *Quarterly Journal of Economics*, 120 (3): 1085-1120.

Fukuyama, F. 1995. *Trust: Social Virtues and the Creation of Prosperity*. New York: The Free Press.

Funder, D.C. 1991. Global Traits: A Neo-Allportian Approach to Personality. *Psychological Science*, 2 (1): 31-39.

Furnham, A. 1984. The Protestant work ethic: A review of the psychological literature. *European Journal of Social Psychology*, 14 (1): 87-104.

Furnham, A. 1990. *The Protestant Work Ethic: The Psychology of Work-Related Beliefs and Behaviours*. London: Routledge.

G

Gabaix, X., and D. Laibson. 2008. The Seven Properties of Good Models. In: Caplin, A., and A. Schotter, Eds. *The Methodologies of Modern Economics: Foundations of Positive and Normative Economics*. New York: Oxford University Press.

Gächter, S. B. Herrmann, and C. Thöni. 2010. Culture and Cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365 (1553): 2651-2661.

Gangestad, S.W., M.G. Haselton, and D.M. Buss. 2006. Evolutionary foundations of cultural variation: evoked culture and mate preferences. *Psychological Inquiry*, 17 (2): 75-95.

Ganzeboom, H.B.G., and D.J. Treiman. 1996. Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25 (3): 201-239.

Garretsen, H., R. Lensink, and E. Sterken. 2004. Growth, Financial Development, Societal Norms, and Legal Institutions. *Journal of International Financial Markets, Institutions & Money*, 14 (2): 165-183.

Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Giddens, A. 1992. *Introduction to Weber, M. 1992. The Protestant Ethic and the Spirit of Capitalism* [Translated by T. Parsons, 1930]. London and New York: Routledge.

Gilleard, C.J. 1989. The achieving society revisited: A further analysis of the relation between national economic growth and need achievement. *Journal of Economic Psychology*, 10 (1): 21-34.

Gintis, H., and P. Romer. 1998. *The Human Side of Economic Analysis: Economic Environments and the Evolution of Norms and Preferences*. Unpublished manuscript, University of Massachusetts, Amherst and University of California, Berkeley.

Glaeser, E., G. Ponzetto, and A. Shleifer. 2007. Why does democracy need education? *Journal of Economic Growth,* 12 (2): 77-99.

Glahe, F., and F. Vorhies. 1989. Religion, liberty and economic development: An empirical investigation. *Public Choice,* 62 (3): 201-215.

Glimcher, P.W., and A. Rustichini. 2004. Neuroeconomics: The Consilience of Brain and Decision. *Science*, 306 (15 October): 447-452.

Gneezy, U., K.L. Leonard, and J.A. List. 2009. Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77 (5): 1637-1664.

Gorodnichenko, Y., and G. Roland. 2010. *Culture, Institutions and the Wealth of Nations*. NBER Working Paper No. 16368. Cambridge, MA: National Bureau of Economic Research.

Granato, J., R. Inglehart, and D. Leblang. 1996. The Effect of Cultural Values on Economic Development: Theory, Hypotheses, and Some Empirical Tests. *American Journal of Political Science*, 40 (3): 607-631.

Greif, A. 1994. Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, 102 (5): 912-950.

Greif, A. 2006. *Institutions: Theory and History. Comparative and Historical Institutional Analysis*. Cambridge: Cambridge University Press.

Grether, D.M., and C.R. Plott. 1979. Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review*, 69 (4): 623-338.

Gruber, J.H., and S. Mullainathan. 2005. Do Cigarette Taxes Make Smokers Happier. *Advances in Economic Analysis & Policy*, 5 (1): Article 4. Available at http://www.bepress.com/bejeap/advances/vol5/iss1/art4.

Guala, F. 2008. Paradigmatic Experiments: The Ultimatum Game from Testing to Measurement Device. *Philosophy of Science*, 75 (5): 658-669.

Guiso, L., P. Sapienza, and L. Zingales. 2003. People's opium? Religion and economic attitudes. *Journal of Monetary Economics*, 50 (1): 225-282.

Guiso, L., P. Sapienza, and L. Zingales. 2006. Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, 20 (1): 23-48.

Gulliksen, H. 1950. *Theory of mental tests*. New York: Wiley.

H

Hackney, C.H., and G.S. Sanders. 2003. Religiosity and Mental Health: A Meta-Analysis of Recent Studies. *Journal for the Scientific Study of Religion*, 42 (1): 43-55.

Hadenius, A. 1992. *Democracy and development*. Cambridge: Cambridge University Press.

Häder, M., and S. Häder. 1995. *Turbulenzen im Transformationprozess: Die individuelle Bewältigung des sozialen Wandels in Ostdeutschland 1990-1992*. Opladen: Westdeutscher Verlag.

Haley, K., and D. Fessler. 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26 (3): 245-56.

Hall, J.B., and U. Ludwig. 2006. Economic convergence across German regions in light of empirical findings. *Cambridge Journal of Economics*, 30 (6): 941-953.

Hall, P.A., and D.W. Soskice. 2001. *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.

Haller, M. 2002. Theory and Method in the Comparative Study of Values: Critique and Alternative to Inglehart. *European Sociological Review*, 18 (2): 139-158.

Hames, R. 2001. Human Behavioral Ecology. In: Smelser, N.J., and P.B. Baltes, Eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier.

Hanemann, W.M. 1994. Valuing the Environment through Contingent Valuation. *Journal of Economic Perspectives*, 8 (4): 19-43.

Hanson, N.R. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

Haring, M.J., M.A. Okun, and W.A. Stock. 1984. A quantitative synthesis of literature on work status and subjective well-being. *Journal of Vocational Behavior*, 25 (3): 316-324.

Harré, R. 1998. *The Singular Self: An Introduction to the Psychology of Personhood*. London: Sage.

Harrisson, L.E. 1992. *Who Prospers? How Cultural Values Shape Economic and Political Success*. New York: Basic Books.

Harrison, L.E., and S.P. Huntington, Eds. 2000. *Culture Matters: How Values Shape Human Progress*. New York: Basic Books.

Hayes, J., and P. Nutman. 1981. *Understanding the Unemployed: The psychological effects of unemployment*. London: Tavistock Publications.

Headey, B., and A. Wearing. 1992. *Understanding Happiness: A theory of subjective well-being*. Melbourne: Longman Cheshire.

Hechter, M., H. Kim, and J. Baer. 2005. Prediction Versus Explanation in the Measurement of Values. *European Sociological Review*, 21 (2): 91-108.

Helliwell, J.F. 2006. Well-Being, Social Capital and Public Policy: What's New? *Economic Journal*, 116 (510): C34-C45.

Helliwell, J.F., and C.P. Barrington-Leigh. 2010a. *Measuring and Understanding Subjective Well-Being*. NBER Working Paper No. 15887. Cambridge, MA: National Bureau of Economic Research.

Helliwell, J.F., and C.P. Barrington-Leigh. 2010b. *How Much is Social Capital Worth?* NBER Working Paper No. 16025. Cambridge, MA: National Bureau of Economic Research.

Helliwell, J.F., and H. Huang, 2008. How's Your Government? International Evidence Linking Good Government and Well-Being. *British Journal of Political Science*, 38 (4): 595-619.

Hempel, C.G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, C.G., and P. Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science*, 15 (2): 135-175.

Henrich, J. 2000. Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 90 (4): 973-979.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, Eds. 2004. *Foundations of Human Sociality. Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review, Papers and Proceedings*, 91 (2): 73-78.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath. M. Alvard, A. Barr, J. Ensminger, N. Smith Henrich, K. Hill, F. Gil-White, M. Gurven, F.W. Marlowe, J.Q. Patton, and D. Tracer. 2005. Economic man in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28 (6): 795-855.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J.C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. 2006. Costly Punishment Across Human Societies. *Science*, 312 (26 June): 1767-1770.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J.C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. 2010. Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science*, 327 (19 March): 1480-1484.

Herrmann, B., C. Thoni, and S. Gächter. 2008. Antisocial punishment across societies. *Science*, 319 (7 March): 1362-1367.

Heukelom, F. 2009a. *Kahneman and Tversky and the Making of Behavioral Economics. Economics*. PhD Thesis, Amsterdam: University of Amsterdam.

Heukelom, F. 2009b. *Origin and Interpretation of Internal and External Validity in Economics*. Nijmegen Center for Economics (NiCE) Working Paper, No. 09-111.

Hirschman, A.O. 1984. Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse. *American Economic Review, Papers and Proceedings*, 74 (2): 89-96.

Hodgson, G.M. 2008. Review Essay: Prospects for Economic Sociology. *Philosophy of the Social Sciences*, 38 (1): 133-149.

Hoffman, E., K. McCabe, J. Shachat, and V. Smith. 1994. Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7 (3): 346-380.

Hoffman, S.R. 2004. Islam and democracy. *Comparative Political Studies*, 37 (6): 652-676.

Hofstede, G. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.

Hofstede, G. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Second edition. Thousand Oaks, CA: Sage.

Hofstede, G. 2002. Dimensions do not exist: A reply to Brendan McSweeney. *Human Relations*, 55 (11): 1355-1361.

Hofstede, G. 2006. What did GLOBE really measure? Researchers' minds versus respondents' minds. *Journal of International Business Studies*, 37 (6): 882-896.

Hofstede, G. 2010. The GLOBE debate: Back to relevance. *Journal of International Business Studies*, 41 (8): 1339-1346.

Holländer, H. 2001. On the validity of utility statements: Standard theory versus Duesenberry's. *Journal of Economic Behavior & Organization*, 45 (3): 227-249.

Holt, C.A., and S.K. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review*, 92 (5): 1644-1655.

Hopkins, K.D. 1998. *Educational and psychological measurement and evaluation*. Eighth edition. Boston: Allyn & Bacon.

House, R.J., P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta. 2004. *Culture, Leadership and Organizations: The Globe Study of 62 Societies*. Thousand Oaks, CA: Sage Publications.

Hui, M.K., K. Au, and H. Fock. 2004. Empowerment effects across cultures. *Journal of International Business Studies*, 35 (1): 46-60.

Huntington, S.P. 1984. Will more countries be democratic? *Political Science Quarterly*, 99 (2): 193-218.

Huntington, S.P. 1993. The clash of civilizations. *Foreign Affairs*, 72 (3): 22-50.

Huntington, S.P. 1996. *The clash of civilizations and the remaking of the World order*. New York: Simon & Schuster.

Husted, B.W. 1999. Wealth, Culture, and Corruption. *Journal of International Business Studies*, 30 (2): 339-359.

I

Iannaccone, L.R. 1988. A formal model of Church and sect. *American Journal of Sociology*, 94 (S1): S241-S268.

Iannaccone, L.R. 1992a. Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives. *Journal of Political Economy*, 100 (2): 271-291.

Iannaccone, L.R. 1992b. Religious markets and the economics of religion. *Social Compass*, 39 (1): 123-131.

Iannaccone, L.R. 1998. Introduction to the Economics of Religion. *Journal of Economic Literature*, 36 (3): 1465-1495.

Iannaccone, L.R., and E. Berman. 2006. Religious extremism: The good, the bad, and the deadly. *Public Choice*, 128 (1-2): 109-129.

Ichino, A., and G. Maggi. 2000. Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm. *Quarterly Journal of Economics*, 115 (3): 1057-1090.

Inglehart, R. 1987. Value Change in Industrial Societies. *American Political Science Review*, 81 (4): 1289-1319 (Controversy with S.C. Flanagan).

Inglehart, R. 1988. The Renaissance of Political Culture. *American Political Science Review*, 82 (4): 1203-1230.

Inglehart, R. 1990. *Culture Shift in Advanced Industrial Society*. Princeton, NJ: Princeton University Press.

Inglehart, R. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton, NJ: Princeton University Press.

Inglehart, R., and W.E. Baker. 2000. Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, 65 (1): 19-51.

Inglehart, R., and P. Norris. 2003. The true clash of civilizations. *Foreign Policy*, 135 (March-April): 63-70.

Inglehart, R., P. Norris, and C. Welzel. 2002. Gender equality and democracy. *Comparative Sociology,* 1 (3-4): 321-345.

Inglehart, R., R. Foa, C. Peterson, and C. Welzel. 2008. Development, Freedom, and Rising Happiness A Global Perspective (1981-2007). *Perspectives on Psychological Science*, 3 (4): 264-285.

Isaac, R.M., and D. James. 2000. Just who are you calling risk averse? *Journal of Risk and Uncertainty*, 20 (2): 177-187.

J

Jackman, R.W., and R.A. Miller. 1996. A Renaissance of Political Culture? *American Journal of Political Science*, 40 (3): 632-659.

Jamal, A., and M. Tessler. 2008. The Democracy Barometers: Attitudes in the Arab world. *Journal of Democracy*, 19 (1): 97-110.

Javidan, M., R.J. House, P.W. Dorfman, P.J., Hanges, and M. Sully de Luque. 2006. Conceptualizing and measuring cultures and their consequences: A comparative review of GLOBE's and Hofstede's approaches. *Journal of International Business Studies*, 37 (6): 897-914.

Jianakoplos, N.A., and A. Bernasek. 1998. Are Women More Risk Averse? *Economic Inquiry*, 36 (4): 620-630.

Johansson-Stenman, O., F. Carlsson, and D. Daruvala. 2002. Measuring future grandparents' preferences for equality and relative standing. *Economic Journal*, 112 (479): 362-383.

Jones, H.B. 1997. The Protestant work ethic: Weber's model and the empirical literature. *Human Relations*, 50 (7): 757-778.

Jowell, R., and the Central Co-ordinating Team. 2007. *European Social Survey 2006/2007: Technical Report*. London: Centre for Comparative Social Surveys, City University. Data archived by the Norwegian Social Science Data Services (NSD).

K

Kahn, H. 1979. *World Economic Development: 1979 and Beyond*. Boulder: Westview Press.

Kahneman, D. 1999. Objective Happiness. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Kahneman, D. 2000. Experienced utility and objective happiness: A moment-based approach. In: Kahneman, D., and A. Tversky, Eds. *Choices, Values and Frames*. New York: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D., E. Diener, and N. Schwarz, Eds. 1999. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Kahneman, D., and A.B. Krueger. 2006. Developments in the Measurement of Subjective Well-Being. *Journal of Economic Perspectives*, 20 (1): 3-24.

Kahneman, D., A.B. Krueger, D. Schkade, N. Schwarz, and A. Stone. 2004a. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306 (3 December): 1776-1780.

Kahneman, D., A.B. Krueger, D. Schkade, N. Schwarz, and A. Stone. 2004b. Toward National Well-Being Accounts. *American Economic Review, Papers and Proceedings*, 94 (2): 429-434.

Kahneman, D., I. Ritov, and D. Schkade. 1999. Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, 19 (1-3): 220-242.

Kahneman, D., and R. Sugden. 2005. Experienced Utility as a Standard of Policy Evaluation. *Environmental & Resource Economics*, 32 (1): 161-181.

Kahneman, D., and R.H. Thaler. 1991. Economic Analysis and the Psychology of Utility: Applications to Compensation Policy. *American Economic Review, Papers and Proceedings*, 81 (2): 341-352.

Kahneman, D., and R.H. Thaler. 2006. Anomalies: Utility Maximization and Experienced Utility. *Journal of Economic Perspectives*, 20 (1): 221-234.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47 (2): 263-292.

Kahneman, D., P. Wakker, and R. Sarin. 1997. Back to Bentham? Explorations of Experienced Utility. *Quarterly Journal of Economics*, 112 (2): 375-405.

Kamens, D. 1988. Education and democracy: A comparative institutional analysis. *Sociology and Education*, 61 (2): 114-127.

Kammann, R., M. Farry, and P. Herbison. 1984. The analysis and measurement of happiness as a sense of well-being. *Social Indicators Research*, 15 (2): 91-115.

Kaplan, S. 1987. Aesthetics, affect, and cognition: Environmental preference from an evolutionary perspective. *Environment and Behavior*, 19 (1): 3-32.

Karatnycky, A. 2002. Muslim countries and the democracy gap. *Journal of Democracy*, 13 (1): 99-112.

Keely, L.C., and C.M. Tan. 2005. Understanding preferences for income redistribution. *Journal of Public Economics*, 92 (5-6): 944-961.

King, L.A., and C.K. Napa. 1998. What Makes a Life Good? *Journal of Personality and Social Psychology*, 75 (1): 156-165.

Kirkman, B.L., K.B. Lowe, and C.B. Gibson. 2006. A quarter century of Culture's Consequences: A review of empirical research incorporating Hofstede's cultural value framework. *Journal of International Business Studies*, 37 (3): 285-320.

Kluckhohn, C.K.M., and H.A. Murray. 1948. *Personality in Nature, Society, and Culture*. New York: Knopf.

Kluckhohn, F.R., and F.L. Strodtbeck. 1961. *Variations in value orientations*. Evanston, Il: Row, Peterson.

Koenig, H.G., and D.B. Larson. 2001. Religion and mental health: Evidence for an association. *International Review of Psychiatry*, 13 (2): 67-78.

Kogut, B., and H. Singh. 1988. The Effect of National Culture on the Choice of Entry Mode. *Journal of International Business Studies*, 19 (3): 411-432.

Koivumaa-Honkanen, H., R. Honkanen, H. Viinamäki, K. Heikkilä, J. Kaprio, and M. Koskenvuo. 2001. Life Satisfaction and Suicide: A 20-Year Follow-Up Study. *American Journal of Psychiatry*, 158 (3): 433-439.

Kolinsky, E., and H.M. Nickel. 2003. *Reinventing gender: Women in Eastern Germany since reunification*. London and Portland: Frank Cass.

Koopmans, T.C. 1947. Measurement without theory. *Review of Economic Statistics*, 29 (3): 161-172.

Kőszegi, B., and M. Rabin. 2007. Mistakes in Choice-Based Welfare Analysis. *American Economic Review, Papers and Proceedings*, 97 (2): 477-481.

Kőszegi, B., and M. Rabin. 2008. Choices, situations, and happiness. *Journal of Public Economics*, 92 (8-9): 1821-1832.

Krueger, A.B. 2008. Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox—Comment. *Brookings Papers on Economic Activity*, 2008/1: 95-100.

Krueger, A.B., and D.A. Schkade. 2008. The reliability of subjective well-being measures. *Journal of Public Economics*, 92 (8-9): 1833-1845.

Kuppens, P., A. Realo, and E. Diener. 2008. The Role of Positive and Negative Emotions in Life Satisfaction Judgment Across Nations. *Journal of Personality and Social Psychology*, 95 (1): 66-75.

Kuran, T. 1995. *Private truths, public lies: The social consequences of preference falsification*. Cambridge: Harvard University Press.

Kuran, T. 2003. The Islamic commercial crisis: Institutional roots of economic underdevelopment in the Middle East. *Journal of Economic History,* 63 (2): 414-447.

Kuran, T. 2009. Explaining the economic trajectories of civilizations: The systemic approach. *Journal of Economic Behavior & Organization*, 71 (3): 593-605.

Kurrild-Klitgaard, P., M.K. Justensen, and R. Klemmensen. 2006. The political economy of freedom, democracy and transnational terrorism. *Public Choice*, 128 (1): 289-315.

Kuznets, S. 1972. Problems in Comparing Recent Growth Rates for Developed and Less Developed Countries. *Economic Development and Cultural Change*, 20 (2): 185-209.

Kwan, V.S.Y., M.H. Bond, and T.M. Singelis. 1997. Pancultural explanations for life-satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology,* 73 (5): 1038-1051.

L

Lal, D. 1998. *Unintended Consequences. The Impact of Factor Endowments, Culture, and Politics on Long-Run Economic Performance*. Cambridge and London: The MIT Press.

Landes, D. 1998. *The Wealth and Poverty of Nations*. New York: Norton.

Larsen, R.J., and B.L. Fredrickson. 1999. Measurement Issues in Emotion Research. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Larsen, R.J., E. Diener, and R.A. Emmons. 1985. An evaluation of subjective well-being measures. *Social Indicators Research*, 17 (1): 1-18.

Layard, R. 1980. Human Satisfactions and Public Policy. *Economic Journal*, 90 (363): 737-750.

Layard, R. 2005. *Happiness: Lessons from a new science*. London: Allen Lane / Penguin Books.

Layard, R. 2010. Measuring Subjective Well-Being. *Science*, 327 (29 January): 534-535.

Layard, R., D. Clark, M. Knapp, and G. Mayraz. 2007. Cost-benefit analysis of psychological therapy. *National Institute Economic Review*, 202 (1): 90-98.

Lehmann, H., and G. Roth, Eds. 1993. *Weber's Protestant Ethic. Origins, evidence, contexts*. Cambridge: German Historical Institute, The Press Syndicate of the University of Cambridge.

Lelkes, O. 2006. Tasting freedom: Happiness, religion and economic transition. *Journal of Economic Behavior & Organization*, 59 (2): 173-194.

Lesch, K.-P., D. Bengel, A. Heils, S.Z. Sabol, B.D. Greenberg, S. Petri, J. Benjamin, C.R. Muller, D.H. Hamer, and D.L. Murphy. 1996. Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region. *Science*, 274 (29 November): 1527-1521.

Letendre, K., C.L. Fincher, and R. Thornhill. 2010. Does infectious disease cause global variation in the frequency of intrastate armed conflict and civil war? *Biological Reviews*, 85 (3): 669-683.

Levesque, J., Y. Joanette, B. Mensour, G. Beaudoin, J.-M. Leroux, P. Bourgouin, and M. Beauregard. 2003. Neural correlates of sad feelings in healthy girls. *Neuroscience*, 121 (3): 545-551.

Leviton, L.C. 2001. External Validity. In: Smelser, N.J., and P.B. Baltes, Eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier.

Levitt, S.D., and J.A. List. 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21 (2): 153-174.

Lewin, K. 1935. *A Dynamic Theory of Personality - Selected Papers*. New York and London: McGraw-Hill.

Lichtenstein, S., and P. Slovic. 1971. Reversals of Preference Between Bids and Choices in Gambling Situations. *Journal of Experimental Psychology*, 89 (1): 46-55.

Lichtenstein, S., and P. Slovic. 1973. Response-induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas. *Journal of Experimental Psychology*, 101 (1): 16-20.

Lipset, S.M. 1959. Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *American Political Science Review*, 53 (1): 69-105.

Little, I.M.D. 1949. A Reformulation of the Theory of Consumers' Behavior. *Oxford Economic Papers*, 1 (1): 90-99.

Loehlin, J.C. 1992. *Genes and Environment in Personality Development*. Newbury Park, CA: Sage.

Loewenstein, G. 1999. Experimental Economics from the Vantage-Point of Behavioural Economics. *Economic Journal*, 109 (453): F25-F34.

Lopez, M., and F.A. Lozano. 2009. The Labor Supply of Immigrants in the United States: The Role of Changing Source Country Characteristics. *American Economic Review, Papers and Proceedings*, 99 (2): 35-40.

Lord, F.M., and M.R. Novick. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Louviere, J.J., and D.A. Hensher. 1983. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research*, 10 (3): 348-361.

Louviere, J.J., and G. Woodworth. 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research*, 20 (4): 350-367.

Lucas, R.E., A.E. Clark, Y. Georgellis, and E. Diener. 2004. Unemployment alters the set-point for life satisfaction. *Psychological Science*, 15 (1): 8-13.

Lucas, R.E., E. Diener, and E. Suh. 1996. Discriminant Validity of Well-Being Measures. *Journal of Personality and Social Psychology*, 71 (3): 616-628.

Luechinger, S. 2009. Valuing Air Quality Using the Life Satisfaction Approach. *Economic Journal*, 119 (536): 482-515.

Luechinger, S. 2010. Life satisfaction and transboundary air pollution. *Economics Letters*, 107 (1): 4-6.

Luechinger, S., and P.A. Raschky. 2009. Valuing flood disasters using the life satisfaction approach. *Journal of Public Economics*, 93 (3-4): 620-633.

Luttmer, E. 2001. Group Loyalty and the Taste for Redistribution. *Journal of Political Economy*, 109 (3): 500-529.

Luttmer, E.F.P. 2005. Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics*, 120 (3): 963-1002.

Lynn R. 1991. *The secret of the miracle economy. Different national attitudes to competitiveness and money*. London: The Social Affairs Unit.

Lyubomirsky, S., L. King, and E. Diener. 2005. The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? *Psychological Bulletin*, 131 (6): 803-855.

M

Machlup, F.P. 1946. Marginal Analysis and Empirical Research. *American Economic Review*, 36 (4): 519-554.

MacIntosh, R. 1998. Global attitude measurement: An assessment of the World Values Survey postmaterialism scale. *American Sociological Review*, 63 (3): 452-464.

MacKerron, G., and S. Mourato. 2009. Life satisfaction and air quality in London. *Ecological Economics*, 68 (5): 1441-1453.

Malka, A., and J.A. Chatman. 2003. Intrinsic and Extrinsic Work Orientations as Moderators of the Effect of Annual Income on Subjective Well-Being: A Longitudinal Study. *Personality and Social Psychology Bulletin*, 29 (6): 737-746.

Manski, C.F. 2004. Measuring Expectations. *Econometrica*, 72 (5): 1329-1376.

Marshall, M.G. 2008. *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2008*. Available from http://www.systemicpeace.org/polity/polity4.htm.

Maseland, R., and A. van Hoorn. 2009. Explaining the Negative Correlation between Values and Practices: A Note to the Hofstede-GLOBE Debate. *Journal of International Business Studies*, 40 (3): 527-532.

Maseland, R., and A. van Hoorn. 2010a. Values and Marginal Preferences in International Business. *Journal of International Business Studies*, 41 (8): 1325-1329.

Maseland, R., and A. van Hoorn. 2010b. Why Muslims Like Democracy Yet Have So Little of It. *Public Choice*, doi:10.1007/s11127-010-9642-5.

Mark, M.M., and C.S. Reichardt. 2001. Internal Validity. In: Smelser, N.J., and P.B. Baltes, Eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier.

Matthews, G., I.J. Deary, and M.C. Whiteman. 2009. *Personality Traits*. Third Edition. Cambridge: Cambridge University Press.

McBride, M. 2001. Relative-income effects on subjective well-being in the cross-section. *Journal of Economic Behavior & Organization*, 45 (3): 251-278.

McCleary, R.M., and R.J. Barro. 2006. Religion and Economy. *Journal of Economic Perspectives*, 20 (1): 49-72.

McClelland, D.C. 1961. *The achieving society*. Princeton: Van Nostrand.

McCloskey, D.N. 1983. The Rhetoric of Economics. *Journal of Economic Literature*, 21 (2): 481-517.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P., Ed. *Frontiers in Econometrics*. New York: Academic Press.

McGue, M., and T.J. Bouchard Jr. 1998. Genetic and Environmental Influences on Human Behavioral Differences. *Annual Review of Neuroscience*, 21: 1-24.

McKee-Ryan, F.M., Z. Song, C.R. Wanberg, and A.J. Kinicki. 2005. Psychological and physical well-being during unemployment: A meta-analytic study. *Journal of Applied Psychology*, 90 (1): 53-76.

McMahon, D.M. 2004. From the happiness of virtue to the virtue of happiness: 400 b.c.-a.d. 1780. *Daedalus*, Spring: 5-17.

McSweeney, B. 2002a. Hofstede's model of national cultural differences and their consequences: A triumph of faith – a failure of analysis. *Human Relations*, 55 (1): 89-118.

McSweeney, B. 2002b. The essentials of scholarship: A reply to Geert Hofstede. *Human Relations*, 55 (11): 1363-1382.

Merton, R.K., D.L. Sills, and S.M. Stigler. 1984. The Kelvin dictum and social science: An excursion into the history of an idea. *Journal of the History of the Behavioral Sciences*, 20 (4): 319-331.

Michael, R.T., and G.S. Becker. 1973. On The New Theory of Consumer Behavior. *Swedish Journal of Economics*, 75 (4): 378-396.

Midlarsky, M.I. 1998. Democracy and Islam: Implications for civilizational conflict and the democratic peace. *International Studies Quarterly*, 42 (3): 485-511.

Minkenberg, M. 2007. Democracy and religion: Theoretical and empirical observations on the relationship between Christianity, Islam and liberal democracy. *Journal of Ethnic and Migration Studies*, 33 (6): 887-909.

Mogahed, D. 2006. *Special report: Muslim world*, *Islam and democracy*. Washington, DC: Gallup Poll Consulting University Press.

Mokyr, J. 2008. The Institutional Origins of the Industrial Revolution. In: Helpman, E., Ed. *Institutions and Economic Performance*. Cambridge and London: Harvard University Press.

Morgenstern, O. 1950. *On the Accuracy of Economic Observations*. Second edition (1963). Princeton, NJ: Princeton University Press.

Morgenstern, O. 1975. *Does GNP Measure Growth and Welfare?* Working Papers, C.V. Starr Center for Applied Economics. New York: New York University.

Morris, C.W. 1956. *Varieties of human value*. Chicago: University of Chicago Press.

Morris, M.H., D.L. Davis, and J.W. Allen. 1994. Fostering corporate entrepreneurship: Cross-cultural comparisons of the importance of individualism versus collectivism. *Journal of International Business Studies*, 25 (1): 65-89.

Moscati, I. 2007. Early Experiments in Consumer Demand Theory: 1930-1970. *History of Political Economy*, 39 (3): 359-401.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32 (3): 385-397

Moulton, B.R. 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics*, 72 (2): 334-338.

Mullainathan, S., and R. Thaler. 2000. *Behavioral Economics*. NBER Working Paper No. 7948. Cambridge, MA: National Bureau of Economic Research.

Murray, D.R., and M. Schaller. 2010. Historical Prevalence of Infectious Diseases Within 230 Geopolitical Regions: A Tool for Investigating Origins of Culture. *Journal of Cross-Cultural Psychology*, 41 (1): 99-108.

N

Navarro, A. 2009. Genoeconomics: Promises and Caveats for a New Field. *Annals of the New York Academy of Sciences*, 1167 (1): 57-65.

Nettle, D. 2005. *Happiness: The science behind your smile*. Oxford: Oxford University Press.

Newman, M.L. and R.A. Josephs. 2009. Testosterone as a Personality Variable. *Journal of Research in Personality*, 43 (2): 258-259.

Ng, Y.-K. 1987. Relative-income effects and the appropriate level of public expenditure. *Oxford Economic Papers*, 39 (3): 293-300.

Noland, M. 2005. Religion and economic performance. *World Development*, 33 (8): 1215-1232.

Norris, P., and R. Inglehart. 2004. *Sacred and Secular*. Cambridge: Cambridge University Press.

North, D.C. 1981. *Structure and Change in Economic History*. New York: Norton.

North, D.C. 1990. *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.

North, D.C. 2005. *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, MA: Belknap Press of Harvard University Press.

Nunnally, J.C. 1967. *Psychometric theory*. New York: McGraw Hill.

O

Oishi, S. 2010. Culture and well-being: Conceptual and methodological issues. In: Diener, E., D. Kahneman, and J.F. Helliwell, Eds. *International differences in well-being*. Oxford: Oxford University Press.

Oishi, S., E. Diener, R.E. Lucas, and E. Suh. 1999. Cross-Cultural Variations in Predictors of Life Satisfaction: Perspectives From Needs and Values. *Personality and Social Psychology Bulletin*, 25 (8): 980-990.

Oishi, S., E. Diener, E. Suh, and R.E. Lucas. 1999. Value as a Moderator in Subjective Well-Being. *Journal of Personality*, 67 (1): 157-184.

Orians, G.H., and J.H. Heerwagen. 1992. Evolved responses to landscapes. In: Barkow, J.H., L. Cosmides, and J. Tooby, Eds. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford and New York: Oxford University Press.

Ortony, A., and T.J. Turner. 1990. What's Basic About Basic Emotions? *Psychological Review*, 97 (3): 315-331.

Oswald, A.J., and N. Powdthavee. 2008. Death, Happiness, and the Calculation of Compensatory Damages. *Journal of Legal Studies*, 37 (2): S217-S251.

Oswald, A.J., and S. Wu. 2010. Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A. *Science*, 327 (29 January): 576-579.

Ott, J.C. 2010a. Good Governance and Happiness in Nations: Technical Quality Precedes Democracy and Quality Beats Size. *Journal of Happiness Studies*, 11 (3): 353-368.

Ott, J. 2010b. Greater Happiness for a Greater Number: Some Non-controversial Options for Governments. *Journal of Happiness Studies*, 11 (5): 631-647.

Ouweneel, P., and R. Veenhoven. 1991. Cross-national differences in happiness: Cultural Bias or Societal Quality? In: N. Bleichrodt, and P.J. Drenth, Eds. *Contemporary issues in cross-cultural psychology: selected papers from a regional conference of the International Association for Cross-cultural Psychology*. Amsterdam: Swets and Zeitlinger.

P

Pargament, K.I. 2002. The Bitter and the Sweet: An Evaluation of the Costs and Benefits of Religiousness. *Psychological Inquiry*, 13 (3): 168-181.

Park, N., and E.S. Huebner. 2005. A Cross-Cultural Study of the Levels and Correlates of Life Satisfaction among Adolescents. *Journal of Cross-Cultural Psychology*, 36 (4): 444-456.

Park, S.H., and G.R. Ungson 1997. The Effect of National Culture, Organizational Complementarity, and Economic Motivation on Joint Venture Dissolution. *Academy of Management Journal*, 40 (2): 279-307.

Pavot, W., and E. Diener. 1993a. Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5 (2): 164-172.

Pavot, W., and E. Diener. 1993b. The Affective and Cognitive Context of Self-Reported Measures of Subjective Well-Being. *Social Indicators Research*, 28 (1): 1-20.

Payne, J.W., J.R. Bettman, and D.A. Schkade. 1999. Measuring constructed preferences. *Journal of Risk and Uncertainty*, 19 (1-3): 243-270.

Phelps, E.S. 2006. *Economic Culture and Economic Performance: What Light Is Shed on the Continent's Problem?* Paper presented at the Conference of CESifo and Center on Capitalism and Society on "Perspectives on the Performance of the Continent's Economies", Venice, 21-22 July 2006.

Phelps, E., and G. Zoega. 2009. Entrepreneurship, culture and openness. In: Audretsch, D.B., R.E. Litan, and R.J. Strom, Eds. *Entrepreneurship and Openness: Theory and Evidenc*e. Cheltenham: Edward Elgar.

Pigou, A.C. 1952. *The economics of welfare*. Fourth edition (1920), reprinted with eight new appendices. London: MacMillan.

Pittau, M.G., R. Zelli, and A. Gelman. 2010. Economic Disparities and Life Satisfaction in European Regions. *Social Indicators Research*, 96 (2): 339-361.

Plomin, R., M.J. Owen, and P. McGuffin. 1994. The Genetic Basis of Complex Human Behaviors. *Science*, 264 (17 June): 1733-1739.

Plott, C.R., and V.L. Smith, Eds. 2008. *Handbook of Experimental Economics Results*. Vol. 1. New York: Elsevier Press.

Powell, M., and D. Ansic. 1997. Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18 (6): 605-628.

Prescott, E.C. 1986. Theory ahead of business-cycle measurement. *Carnegie-Rochester Conference Series on Public Policy*, 25 (1): 11-44.

Pryor, F.L. 2005. National Values and Economic Growth. *American Journal of Economics and Sociology*, 64 (2): 451-83.

Putnam, R. 1993. *Making Democracy Work*. Princeton, NJ: Princeton University Press.

R

Rabin, M. 1998. Psychology and Economics. *Journal of Economic Literature*, 36 (1): 11-46.

Rabin, M. 2002. A perspective on psychology and economics. *European Economic Review*, 46 (4-5): 657-685.

Rainville, P., A, Bechara, N. Naqvi, and A.R. Damasio. 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61 (1): 5-18.

Raudenbush, S.W., and A.S. Bryk. 2002. *Hierarchical Linear Models*. Second Edition. Thousand Oaks, CA: Sage Publications.

Redding, S.J., and D.M. Sturm. 2008. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review*, 98 (5): 1766-1797.

Rehdanz, K., and D. Maddison. 2008. Local Environmental Quality and Life-Satisfaction in Germany. *Ecological Economics*, 64 (4): 787-797.

Robson, A.J. 2001. The Biological Basis of Economic Behavior. *Journal of Economic Literature*, 39 (1): 11-33.

Robson, A.J. 2002. Evolution and Human Nature. *Journal of Economic Perspectives*, 16 (2): 89-106.

Rodgers, W.L., A.R. Herzog, and F.M. Andrews. 1988. Interviewing Older Adults: Validity of Self-Reports of Satisfaction. *Psychology and Aging*, 3 (5): 264-272.

Rokeach, M. 1973. *The Nature of Human Values*. New York: The Free Press.

Rose, R. 2002. How Muslims View Democracy: Evidence from Central Asia. *Journal of Democracy*, 13 (4): 102-111.

Rosen, S. 1996. Public Employment and the Welfare State in Sweden. *Journal of Economic Literature*, 34 (2): 729-740.

Ross, M.L. 2001. Does oil hinder democracy? *World Politics*, 53 (3): 325-361.

Roth, A.E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. 1991. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Revi*ew, 81 (5): 1068-1095.

Rowley, C.K., and N. Smith. 2009. Islam's democracy paradox: Muslims claim to like democracy, so why do they have so little? *Public Choice*, 139 (3): 273-299.

Rubinstein, A. 2006. *Lecture Notes in Microeconomic Theory*. Princeton and Oxford: Princeton University Press.

Rustichini, A. 2005. Neuroeconomics: Present and Future. *Games and Economic Behavior*, 52 (2): 201-212.

Ryff, C.D., B.H. Singer, and G. Dienberg Love. 2004. Positive health: connecting well-being with biology. *Philosophical Transactions Royal Society London B: Biological Sciences*, 359 (1449): 1383-1394.

S

Sagarin, B.J., and J.J. Skowronski. 2009. The implications of imperfect measurement for free-choice carry-over effects: Reply to M. Keith Chen's (2008). ''Rationalization and Cognitive Dissonance. Do Choices Affect or Reflect Preferences?" *Journal of Experimental Social Psychology*, 45 (2): 421-423.

Said, E. 1978. *Orientalism*. New York: Random House.

Salmon, W.C. 1989. Four decades of scientific explanation. In: Kitcher, P., and W.C. Salmon, Eds. *Scientific explanation*. Minneapolis: University of Minnesota Press.

Samuelson, P.A. 1938a. A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5 (17): 61-71.

Samuelson, P.A. 1938b. A Note on the Pure Theory of Consumer's Behaviour: An Addendum. *Economica*, 5 (19): 353-354.

Samuelson, P.A. 1948. Consumption Theory in Terms of Revealed Preferences. *Economica*, 15 (60): 243-253.

Sapienza, P., L. Zingales, and D. Maestripieri. 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, 106 (36): 15268-15273.

Savani, K., H.R. Markus, and A.L. Conner. 2008. Let your preference be your guide? Preference and choices are more tightly linked for North Americans than for Indians. *Journal of Personality and Social Psychology*, 95 (4): 861-876.

Schaller, M. 2006. Parasites, behavioral defenses, and the social psychological mechanisms through which cultures are evoked. *Psychological Inquiry*, 17 (2): 96-101.

Schaller, M., and D.R. Murray. 2008. Pathogens, personality, and culture: disease prevalence predicts worldwide variability in sociosexuality, extraversion, and openness to experience. *Journal of Personality and Social Psychology*, 95 (1): 212-221.

Schaltegger, C.A., and B. Torgler. 2010. Work ethic, Protestantism, and human capital. *Economics Letters*, 107 (2): 99-101.

Schimmack, U., P. Radhakrishnan, S. Oishi, V. Dzokoto, and S. Ahadi. 2002. Culture, Personality, and Subjective Well-Being: Integrating Process Models of Life Satisfaction. *Journal of Personality and Social Psychology*, 82 (4): 582-593.

Schinka, J.A., R.M. Busch, and N. Robichaux-Keene. 2004. A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Molecular Psychiatry*, 9 (2): 197-202.

Schubert, R., M. Gysler, M. Brown, and H.-W. Brachinger. 1999. Financial Decision-Making: Are Women Really More Risk Averse? *American Economic Review, Papers and Proceedings*, 89 (2): 381-385.

Schwartz, S.H., and L. Sagiv. 1995. Identifying Culture-Specifics in the Content and Structure of Values. *Journal of Cross-Cultural Psychology*, 26 (1): 92-116.

Schwartz, S.H. 1994. Beyond individualism/collectivism: New cultural dimensions of values. In: Kim, U., H.C. Triandis, C. Kagitçibasi, S.C. Choi, and G. Yoon, Eds. *Individualism and collectivism: Theory, method, and applications*. Thousand Oaks, CA: Sage.

Schwartz, S., and W. Bilsky. 1987. Towards a psychological structure of human values. *Journal of Personality and Social Psychology*, 53 (3): 550-562.

Schwarz, N., and F. Strack. 1999. Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Scollon, C.N., C. Kim-Prieto, and E. Diener. 2003. Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, 4 (1): 5-34.

Scollon, C.N., E. Diener, S. Oishi, and R. Biswas-Diener. 2004. Emotions across cultures and methods. *Journal of Cross-Cultural Psychology*, 35 (3): 304-326.

Seidlitz, L., and E. Diener. 1993. Memory for Positive Versus Negative Events: Theories for the Differences Between Happy and Unhappy Persons. *Journal of Personality and Social Psychology*, 64 (4): 654-664.

Sellers, J.G., M.R. Mehl, and R.A. Josephs. 2007. Hormones and personality: Testosterone as a marker of individual differences. *Journal of Research in Personality*, 41 (1): 126-138.

Sen, A. 1973. Behaviour and the Concept of Preference. *Economica*, 40 (159): 241-259.

Sen, A. 1976. Real National Income. *Review of Economic Studies*, 43 (1): 19-39.

Sen, A. 1977. Rational Fools: A Critique of the Behavioural Foundations of Economic Theory. *Philosophy and Public Affairs*, 6 (4): 317-344.

Sen, A. 1993. Internal Consistency of Choice. *Econometrica*, 61 (3): 495-521.

Sent, E.-M. 2004. Behavioral Economics: How Psychology Made Its (Limited) Way Back Into Economics. *History of Political Economy*, 36 (4): 735-760.

Sgroi, D., E. Proto, A.J. Oswald, and A. Dobson. 2010. *Priming and the Reliability of Subjective Well-being Measures*. Warwick Economic Research Papers No. 935. Department of Economics, University of Warwick.

Shafiq, M.N. 2010. Do education and income affect support for democracy in Muslim countries? Evidence from the Pew Global Attitudes Project. *Economics of Education Review*, 29 (3): 461-469.

Shields, M., and S. Wheatley Price. 2005. Exploring the economic and social determinants of psychological and psychosocial health. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 168 (3): 513-538.

Shiller, R.J., M. Boycko, and V. Korobov. 1991. Popular Attitudes Toward Free Markets: The Soviet Union and the United States Compared. *American Economic Review*, 81 (3): 385-400.

Shiller, R.J., M. Boycko, and V. Korobov. 1992. Hunting for Homo Sovieticus: Situational versus Attitudinal Factors in Economic Behavior. *Brookings Papers on Economic Activity*, 1992/1: 127-194.

Simon, H.A. 1981. *Sciences of the Artificial*. Second Edition. Cambridge, MA: The MIT Press.

Simon, H.A. 1987. behavioral economics. In: Eatwell, J., M. Millgate, and P. Newman, Eds. *The New Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan.

Sinn, H.-W. 2002. Germany's economic unification: An assessment after ten years. *Review of International Economics*, 10 (1): 113-28.

Sinn, H.-W., and F. Westermann. 2001. *Two Mezzogiornos*. NBER Working Paper No. 8125. Cambridge, MA: National Bureau of Economic Research.

Sirgy, M.J., A.C. Michalos, A.L. Ferriss, R.A. Easterlin, D. Patrick, and W. Pavot. 2006. The Quality-of-Life (QOL) research movement: Past, present, and future. *Social Indicators Research*, 76 (3): 343-466.

Smith, A. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Bantam Classic 2003 edition. New York: Bantam.

Smith, P.B. 2006. When elephants fight, the grass gets trampled: The GLOBE and Hofstede projects. *Journal of International Business Studies*, 37 (6): 915-921.

Smith, V.L. 1989. Theory, Experiment and Economics. *Journal of Economic Perspectives*, 3 (1): 151-169.

Smith, T.B., M.E. McCullough, and J. Poll. 2003. Religiousness and Depression: Evidence for a Main Effect and the Moderating Influence of Stressful Life Events. *Psychological Bulletin*, 129 (4): 614-636.

Smits, N., G.J. Mellenbergh, and H.C.M. Vorst. 2003. *The Measurement Versus Prediction Paradox in the Application of Planned Missingness to Psychological and Educational Tests*. SCHOLAR Working Paper Series WP 33/02.

Snijders, T.A.B., and R.J. Bosker. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Snower, D.J., and C. Merkl. 2006. The Caring Hand that Cripples: The East German Labor Market After Reunification. *American Economic Review, Papers and Proceedings*, 96 (2): 375-382.

Solnick, S.J. 2001. Gender Differences in the Ultimatum Game. *Economic Inquiry*, 39 (2): 189-200.

Solnick, S.J., and D. Hemenway. 1998. Is more always better?: A survey on positional concerns. *Journal of Economic Behavior & Organization*, 37 (3): 373-383.

Solnick, S.J., and D. Hemenway. 2005. Are Positional Concerns Stronger in Some Domains than in Others? *American Economic Review, Papers and Proceedings*, 95 (2): 147-151.

Spence, J.T. 1985. Achievement American style: The rewards and costs of individualism. *American Psychologist*, 40 (12): 1285-1295.

Statistisches Ämter des Bundes und der Länder, 2010. *Bruttoinlandsprodukt, Bruttowertschöpfung in den Ländern und Ost-West-Großraumregionen Deutschlands 1991 bis 2009 Reihe 1, Band 1*. Stuttgart, Germany: Statistische Landesamt Baden-Württemberg.

Steptoe, A., J. Wardle, and M. Marmot. 2005. Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proceedings of the National Academy of Sciences*, 102 (18): 6508-6512.

Stevenson, B., and J. Wolfers. 2008. Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. *Brookings Papers on Economic Activity* 2008/1: 1-87.

Stewart, A.L., R.D. Hays, and J.E. Ware, Jr. 1988. The MOS Short-Form General Health Survey: Reliability and Validity in a Patient Population. *Medical Care*, 26 (7): 724-735.

Stigler, G.J. 1950. The Development of Utility Theory. II. *Journal of Political Economy*, 58 (5): 373-396.

Stigler, G.J. 1965. *Essays in the History of Economics*. Chicago: University of Chicago Press.

Stigler, G.J, and G.S. Becker. 1977. De Gustibus Non Est Disputandum. *American Economic Review*, 67 (2): 76-90.

Stone, A.A., S.S. Shiffman, and M.W. Devries. 1999. Ecological Momentary Assessment. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Stutzer, A., and B.S. Frey. 2010. *Recent Advances in the Economics of Individual Subjective Well-Being*. IZA Discussion Paper No. 4850. Bonn: Institute for the Study of Labor.

Stutzer, A., and R. Lalive. 2004. The role of social work norms in job searching and subjective well-being. *Journal of the European Economic Association*, 2 (4): 696-719.

Suh, E., E. Diener, S. Oishi, and H.C. Triandis. 1998. The Shifting Basis of Life Satisfaction Judgments Across Cultures: Emotions Versus Norms. *Journal of Personality and Social Psychology*, 74 (2): 482-493.

Sumner, L.W. 1995. The Subjectivity of Welfare. *Ethics*, 105 (4): 764-790.

Sundén, A.E., and B.J. Surette. 1998. Gender Differences in the Allocation of Assets in Retirement Savings Plans. *American Economic Review, Papers and Proceedings*, 88 (2): 207-211.

Swank, D. 1996. Culture, Institutions, and Economic Growth: Theory, Recent Evidence, and the Role of Communitarian Polities. *American Journal of Political Science*, 40 (3): 660-679.

Swedberg, R. 1990. *Economics and Sociology: Redefining Their Boundaries: Conversations with Economists and Sociologists*. Princeton, NJ: Princeton University Press.

T

Tabellini, G. 2008. Institutions and culture. *Journal of the European Economic Association*, 6 (2-3): 255-294.

Tabellini, G. 2010. Culture and institutions: Economic development in the regions of Europe. *Journal of the European Economic Association*, 8 (4): 677-716.

Taras, V., J. Rowney, and P. Steel. 2009. Half a Century of Measuring Culture: Approaches, Challenges, Limitations, and Suggestions Based on the Analysis of 121 Instruments for Quantifying Culture. *Journal of International Management*, 15 (4): 357-373.

Taras, V., P. Steel, and B.L. Kirkman. 2010. Negative practice-value correlations in the GLOBE data: Unexpected findings, questionnaire limitations and research directions. *Journal of International Business Studies*, 41 (8): 1330-1338.

Tawney, R.H. 1926. *Religion and the Rise of Capitalism*. New York: Harper and Row.

The Conference Board and Groningen Growth and Development Centre. 2008. *Total Economy Database*. January 2008.

Thomas, A.S., and S.L. Mueller. 2000. A case for comparative entrepreneurship: Assessing the relevance of culture. *Journal of International Business Studies*, 31 (2): 287-301.

Thompson, B., and T. Vacha-Haase. 2000. Psychometrics Is Datametrics: The Test Is Not Reliable. *Educational and Psychological Measurement*, 60 (2): 174-195.

Thomson, W. 1889. *Popular Lectures and Addresses*. Vol. I. Constitution of Matter. London and New York: MacMillan.

Thorndike, E.L. 1918. The nature, purposes, and general methods of measurements of educational products. In: Whipple, G.M., Ed. *The Seventeenth yearbook of the National Society for Study of Education*. Part II. The Measurement of Educational Products. Bloomington, IL: Public School Publishing Co.

Thorndike, R.M. 2004. *Measurement and Evaluation in Psychology and Education*. Seventh Edition. addle River, NJ: Prentice Hall.

Tooby, J., and L. Cosmides. 1992. The Psychological Foundations of Culture. In: Barkow, J.H., L. Cosmides, and J. Tooby, Eds. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford and New York: Oxford University Press.

Torgler, B., and C.A. Schaltegger. 2009. Religion and economy: A comment. *Economics Bulletin*, 29 (3): 2227-2230.

Tversky, A., and D. Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science*, 211 (30 January): 453-458.

Tversky, A., and R.H. Thaler. 1990. Preference Reversals. *Journal of Economic Perspectives*, 4 (1): 193-205.

U

Uchida, Y., V. Norasakkunkit, and S. Kitayama. 2004. Cultural construction of happiness: Theory and empirical evidence. *Journal of Happiness Studies*, 5 (3): 223-239.

Uhlaner, L., and R. Thurik. 2007. Postmaterialism influencing total entrepreneurial activity across nations. *Journal of Evolutionary Economics*, 17 (2): 161-185.

Uhlig, H. 2006. Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification. *American Economic Review, Papers and Proceedings*, 96 (2): 383-387.

Uhlig, H. 2008. The slow decline of East Germany. *Journal of Comparative Economics*, 36 (4): 517-541.

Urry, H.L., J.B. Nitschke, I. Dolski, D.C. Jackson, K.M. Dalton, C.J. Mueller, M.A. Rosenkranz, C.D. Ryff, B.H. Singer, and R.J. Davidson. 2004. Making a Life Worth Living: Neural Correlates of Well-Being. *Psychological Science*, 15 (6): 367-372.

V

van de Stadt, H., A. Kapteyn, and S. van de Geer. 1985. The Relativity of Utility: Evidence from Panel Data. *Review of Economics and Statistics*, 67 (2): 179-187.

van de Vijver, F.J.R., and K. Leung. 1997. *Methods and data analysis for cross-cultural research*. Newbury Park: Sage.

van Hoorn, A. 2008. A Short Introduction to Subjective Well-being: Measurement, Correlates and Policy Uses. In: OECD, Ed. *Statistics, Knowledge and Policy 2007: Measuring and Fostering the Progress of Societies*. Paris: OECD Publishing.

van Hoorn, A., R. Mabsout, and E.-M. Sent. 2010. Happiness and Capability: Introduction to the Symposium. *Journal of Socio-Economics*, 39 (3): 339-343.

van Hoorn, A., and R. Maseland. 2010. Cultural Differences Between East and West Germany After 1991: Communist Values versus Economic Performance? *Journal of Economic Behavior & Organization*, 76 (3): 791-804.

van Praag, B.M.S., and B. Baarsma. 2005. Using Happiness Surveys to Value Intangibles: The Case of Airport Noise. *Economic Journal*, 115 (500): 224-246.

van Praag, B.M.S., and P. Frijters. 1999. The Measurement of Welfare and Well-Being: The Leyden Approach. In: Kahneman, D., E. Diener, and N. Schwarz, Eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

van Praag, B.M.S., and N.L. van der Sar. 1988. Empirical Uses of Subjective Measures Of Well-Being. *Journal of Human Resources*, 23 (2): 193-210.

VanderZee, K.I., R. Sanderman, and J. Heyink. 1996. A comparison of two multidimensional measures of health status: The Nottingham Health Profile and the RAND 36-Item Health Survey 1.0. *Quality of Life Research*, 5 (1): 165-174.

Varian, H.R. 2006. Revealed Preference. In: Szenberg, M., L. Ramrattan, and A.A. Gottesman, Eds. *Samuelsonian Economics and the Twenty-First Century*. Oxford: Oxford University Press.

Veenhoven, R. 1984. *Conditions of happiness*. Dordrecht: D. Reidel Publishing Company.

Veenhoven, R. 1987. Cultural bias in ratings of perceived life quality. *Social Indicators Research* 19 (3): 329-334.

Veenhoven, R. 1988. Utility of happiness. *Social Indicators Research*, 20 (4): 333-354.

Veenhoven, R. 1996. Developments in satisfaction research. *Social Indicators Research*, 37 (1): 1-46.

Veenhoven, R. 2002. Why social policy needs subjective indicators. *Social Indicators Research*, 58 (1): 33-45.

Veenhoven, R. 2010a. Greater Happiness for a Greater Number: Is that Possible and Desirable? *Journal of Happiness Studies*, 11 (5): 605-629.

Veenhoven, R. 2010b. How universal is happiness? In: Diener, E., D. Kahneman, and J.F. Helliwell, Eds. *International differences in well-being*. Oxford: Oxford University Press.

Venaik, S., and P. Brewer. 2010. Avoiding uncertainty in Hofstede and GLOBE. *Journal of International Business Studies*, 41 (8): 1294-1315.

Vittersø, J., R. Biswas-Diener, and E. Diener. 2005. The divergent meanings of life satisfaction: Item response modeling of the satisfaction with life scale in Greenland and Norway. *Social Indicators Research*, 74 (2): 327-348.

von Gaudecker, H.-M., A. van Soest, and E. Wengström. 2010. Heterogeneity in Risky Choice Behaviour in a Broad Population. *American Economic Review*, Forthcoming.

W

Wagner, W. 1999. Deutscher, proletarischer und moralischer—Unterschiede zwischen Ost- und Westdeutschland und ihre Erklärung. In: Brähler, E., and H. Berth, Eds. *Deutsch-deutsche Vergleiche - Psychologische Untersuchungen 10 Jahre nach dem Mauerfall*. Berlin: Verlag für Wissenschaft und Forschung.

Wagner, G.G., J.R. Frick, and J. Schupp. 2007. The German socio-economic panel study (SOEP) – scope, evolution and enhancements. *Schmollers Jahrbuch*, 127 (1): 139-169.

Walsh, P. 2005. well-being. In: *The Oxford Companion to Philosophy*. Oxford University Press: Oxford Reference Online.

Watson, D., and L.A. Clark. 1991. Self Versus Peer Ratings of Specific Emotional Traits: Evidence of Convergent and Discriminant Validity. *Journal of Personality and Social Psychology*, 60 (6): 927-940.

Watson, D., L.A. Clark, and A. Tellegen. 1988. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54 (6): 1063-1070.

Way, B.M., and M.D. Lieberman. 2010. Is there a genetic contribution to cultural differences? Collectivism, individualism and genetic markers of social sensitivity. *Social Cognitive & Affective Neuroscience*, 5 (2-3): 203-211.

Weber, M. 1904/1905 [1930]. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons. London and New York: Routledge.

Weil, D.N. 2009. *Economic Growth*. Second edition. Boston: Pearson Addison Wesley.

Weiner, M. 1987. Empirical democratic theory. In: Weiner, M., and E. Özbudun, Eds. *Competitive elections in developing countries*. Durham, NC: Duke University Press.

Weimann, J. 1994. Individual behaviour in a free riding experiment. *Journal of Public Economics*, 54 (2): 185-200.

Wennekers, A., R. Thurik, A. van Stel, and N. Noorderhaven. 2007. Uncertainty avoidance and the rate of business ownership across 21 OECD countries, 1976-2004. *Journal of Evolutionary Economics*, 17 (2): 133-160.

Welsch, H. 2002. Preferences over Prosperity and Pollution: Environmental Valuation Based on Happiness Surveys. *Kyklos*, 55 (4): 473-494.

Welsch, H. 2006. Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics*, 58 (4): 801-813.

Welsch, H. 2007. Environmental welfare analysis: A life satisfaction approach. *Ecological Economics*, 62 (3-4): 544-551.

Welsch, H. 2009. Implications of happiness research for environmental economics. *Ecological Economics*, 68 (11): 2735-2742.

Welsch, H., and J. Kühling. 2009. Using Happiness Data for Environmental Valuation: Issues and Applications. *Journal of Economic Surveys*, 23 (2): 385-406.

Whiteley, P., H.D. Clarke, D. Sanders, and M.C. Stewart. 2010. Government Performance and Life Satisfaction in Contemporary Britain. *Journal of Politics*, 72 (3): 733-746.

Wierzbicka, A. 1999. *Emotions Across Languages and Cultures: Diversity and Universals*. Paris: Cambridge University Press.

Wierzbicka, A. 2004. 'Happiness' in cross-linguistic & cross-cultural perspective. *Daedalus*, Spring: 34-43.

Winkelmann, L., and R. Winkelmann. 1998. Why are the unemployed so unhappy? Evidence from panel data. *Economica*, 65 (257): 1-15.

Wilkinson, R.G., and K.E. Pickett. 2006. Income inequality and population health: A review and explanation of the evidence. *Social Science & Medicine*, 62 (7): 1768-1784.

Williamson, C.R. 2009. Informal institutions rule: Institutional arrangements and economic performance. *Public Choice*, 139 (3): 371-387.

Williamson, C.R., and R.L. Mathers 2010. Economic freedom, culture, and growth. *Public Choice*, doi:10.1007/s11127-010-9656-z.

Williamson, O. 1985. *The Economic Institutions of Capitalism. Firms, Markets, Relational Contracting*. New York: Free Press.

Williamson, O. 2000. The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, 38 (3): 595-613.

Wilson, E.O. 1975. *Sociobiology: The New Synthesis*. Twenty-fifth Anniversary Edition (2000). Cambridge, MA: Harvard University Press.

Wilson, W. 1967. Correlates of avowed happiness. *Psychological Bulletin*, 67 (4): 294-306.

Winterhalder, B., and E. Smith. 2000. Analyzing Adaptive Strategies: Human Behavioral Ecology at Twenty-Five. *Evolutionary Anthropology*, 9 (2): 51-72.

Wooldridge, J.M. 2003. Cluster-Sample Methods in Applied Econometrics. *American Economic Review*, *Papers and Proceedings*, 93 (2): 133-138.

World Bank. 2010. *World development indicators online database, 2009*. Washington, DC: World Bank.

World Values Survey Association. 2009. *World Values Survey 1981-2008 Official Aggregate v.20090901*. Aggregate File Producer: ASEP/JDS, Madrid.

Z

Zimmermann, A.C., and R.A. Easterlin. 2006. Happily ever after? Cohabitation, marriage, divorce, and happiness in Germany. *Population and Development Review*, 32 (3): 511-528.

Summary (in Dutch)

De Economie van Heterogene Voorkeuren

Ieder mens is in sommige opzichten uniek, in andere opzichten precies hetzelfde als ieder ander, en in weer andere opzichten precies hetzelfde als *sommige* anderen. Dit proefschrift gaat over verschillen tussen groepen mensen, specifiek in hun voorkeuren zoals de voorkeur voor werk, gelijkheid of inkomen.

Traditioneel besteedt de economische wetenschap weinig aandacht aan de voorkeuren of nutsfunctie van mensen. Economen worden niet geacht deze te bestuderen maar ze als gegeven en constant te beschouwen. Vanuit methodologisch oogpunt zijn er ook goede redenen om aan te nemen dat de voorkeuren van alle mensen identiek zijn en bovendien niet veranderen. Er lijkt echter sprake van een kentering. Meest prominente voorbeeld is de oprichting in 1997 van het *Norms and Preferences Network* door een groep economen en antropologen uit de Verenigde Staten. Enkele leden van dit tijdelijke onderzoeksnetwerk waren George Akerlof, Colin Camerer, Ernst Fehr, Edward Glaeser, Avner Greif, George Loewenstein, Casey Mulligan en Paul Romer. Het netwerk werd voorgezeten door econoom Herbert Gintis en antropoloog Robert Boyd, en ook Daniel Kahneman, psycholoog en winnaar van de Nobelprijs voor economie, was lid.

De interesse in voorkeuren heeft raakvlakken met meerdere onderzoeksgebieden, zowel binnen als buiten de economische wetenschap. Binnen de economische wetenschap beklemtonen nieuw-institutioneel economen zoals Douglass North en Oliver Williamson het belang van culturele waarden en normen. De idee is dat economische transacties plaats vinden in een maatschappelijke context en dat de allocatie van schaarse middelen niet goed bestudeerd kan worden zonder ook aandacht te besteden aan de inbedding hiervan in sociale structuren. Een tweede vakgebied binnen de economische wetenschap dat raakvlakken heeft met de studie van de economie van heterogene voorkeuren is gedragseconomie. Gedragseconomie richt zich op het verbeteren van economische theorie door deze psychologisch realistischer te maken. Gedragseconomisch onderzoek kijkt daarvoor vaak naar de nutsfunctie van mensen. Een belangrijke bevinding is dat mensen niet alleen egoïstisch zijn zoals vaak wordt aangenomen maar ook altruïstisch. De gebruikte methoden en technieken zijn veelal overgenomen van de psychologie en betreffen in het bijzonder het

gebruik van gedragsexperimenten. Een laatste verwant vakgebied is cross-culturele psychologie. Deze tak van de psychologie kent een lange traditie in het bestuderen van voorkeuren, waarden en culturele verschillen daarin. Zeer bekend is het werk van de Nederlander Geert Hofstede. Het door hem ontwikkelde raamwerk ordent culturele verschillen tussen landen in vijf dimensies en zijn empirische maatstaven worden veel toegepast in wetenschappelijk onderzoek dat de culturele basis voor bepaalde economische en sociale fenomenen bestudeert. In de economische wetenschap worden Hofstede's maatstaven vooralsnog niet veel toegepast en is de *World Values Survey* populairder. Deze grootschalige enquête heeft ongeveer 350.000 individuen uit meer dan 95 verschillende samenlevingen geïnterviewd over hun normen en waarden. Daarnaast bevat de World Values Survey ook gegevens over de persoonlijke achtergronden van respondenten. Economen hebben deze data onder andere gebruikt om te kijken naar religieuze verschillen in economische voorkeuren, bijvoorbeeld de waarde die men hecht aan werk of de houding ten aanzien van inkomensherverdeling.

Het bestuderen van de economie van heterogene voorkeuren staat of valt met het meten van voorkeuren en de groepsverschillen daarin. Het is ook juist omdat er twijfels bestonden over de mogelijkheid om voorkeuren op zinvolle wijze te meten, dat economen lang hebben afgezien van het bestuderen van de economie van heterogene voorkeuren. Alleen als de heterogeniteit van voorkeuren voldoende nauwkeurig en betrouwbaar gemeten wordt, kan het als vertrekpunt dienen voor een beter begrip van economische en sociale fenomenen. Hoe zit het bijvoorbeeld met het effect van Communisme op de voorkeuren van mensen en wat zijn de gevolgen daarvan voor economische uitkomsten zoals werkloosheid en arbeidsinkomen?

Na het inleidende hoofdstuk bestaat dit proefschrift uit vier delen. In de eerste twee delen staat de methode centraal, hoe kunnen we inhoud geven aan heterogene voorkeuren op een manier die tot bruikbare maatstaven van de verschillen in voorkeuren leidt? Deze delen leggen hiermee basis voor het derde deel dat kijkt naar feitelijke verschillen in voorkeuren van specifieke groepen mensen. Het vierde deel bevat de conclusie.

Deel I (Hoofdstuk 2 en 3) evalueert drie gangbare methoden om voorkeuren te meten alsmede groepsverschillen daarin. Uitgangspunt is een intuïtief begrip van voorkeuren als de waardering die mensen hebben voor bepaalde zaken en de criteria die ze hanteren om de wenselijkheid van een situatie te evalueren. De drie methoden zijn:

(1) Het observeren van feitelijk keuzes van mensen, bijvoorbeeld in specifieke gedragsexperimenten (zie boven). Deze methode staat bekend onder de naam *revealed preferences*.

(2) Enquêtes die mensen vragen naar hun voorkeuren, bekend onder de naam *stated preferences*. Voorbeelden zijn de reeds genoemde World Values Survey en het werk van Geert Hofstede.

(3) Het schatten van empirische relaties die laten zien hoeveel bepaalde omstandigheden bijdragen aan het ervaren nut of subjectief welzijn van mensen. Dit kunnen we *experienced preferences* noemen en deze methode staat centraal in dit proefschrift.

Hoofdstuk 2 van de thesis gaat in op de theoretische basis van deze drie methoden en bespreekt bovendien enkele prototypische empirische toepassingen. De kern van dit hoofdstuk is echter de evaluatie van de bruikbaarheid van de drie methoden voor het bestuderen van de economie van heterogene voorkeuren. Hiervoor hanteren we twee criteria, te weten psychometrische kwaliteit en toepasbaarheid. Psychometrische kwaliteit betreft de validiteit of nauwkeurigheid en de precisie of betrouwbaarheid van de maatstaven van verschillen in voorkeuren. Toepasbaarheid is opgesplitst in subcriteria zoals beschikbaarheid en de capaciteit om te verklaren. Aangaande revealed preferences zijn de belangrijkste bevindingen dat er weinig data beschikbaar zijn en dat zij op wetenschapsfilosofische gronden niet goed in staat zijn om economische verschillen tussen groepen te verklaren. Dit maakt revealed preferences weinig toepasbaar voor de studie van de economie van heterogene voorkeuren, tenzij onder specifieke omstandigheden. Aangaande stated preferences data blijkt dat zij mogelijk geplaagd worden door een fundamenteel meetprobleem en daardoor over onvoldoende psychometrische kwaliteit beschikken. De experienced preferences methode is relatief veelbelovend maar tegelijkertijd nog vrij onbekend. Deze onbekendheid betreft vooral de kwaliteit van dit soort data.

Hoofdstuk 3 geeft een nadere analyse van de psychometrische kwaliteit van stated preferences data. Dit gebeurt aan de hand van een paradox in de literatuur die democratische attitudes onderzoekt, namelijk dat men in Moslimlanden erg positief denkt over democratie terwijl deze landen tegelijkertijd relatief weinig democratisch zijn. Een mogelijk verklaring is dat de stated preferences methode geen diepe culturele waarden meet, maar zogenaamde marginale voorkeuren. Stated preferences maatstaven meten dan dus de waarde die men hecht aan meer van een bepaald "goed" bovenop het huidige consumptieniveau. Een eerste ijsje is lekker, een tweede ook nog, maar het tiende niet meer. Toegepast op de democratieparadox betekent dit dat voor inwoners van democratische landen verder aan hun

behoefte aan democratie voldaan is dan voor inwoners van autocratische regimes en dat de eerste groep mensen juist daardoor minder waarde hecht aan nog meer democratie. Voor mensen in weinig democratische omstandigheden geldt het tegenovergestelde en zij hechten veel waarde aan democratie juist omdat zij in het dagelijks leven een gebrek aan democratie ervaren. De empirische analyse in Hoofdstuk 3 laat exact het hierboven beschreven patroon zien, waardoor er dus geen sprake lijkt te zijn van een specifieke Islamitische democratieparadox: de schijnbare democratieparadox van Moslimlanden is een uitdrukking van een veel algemener patroon. Belangrijker voor de beoordeling van de bruikbaarheid van de stated preferences methode is dat deze methode inderdaad een meetprobleem lijkt te hebben: stated preferences maatstaven van (heterogene) voorkeuren meten geen fundamentele waarden maar marginale voorkeuren, die op hun beurt bepaald worden door omstandigheden, dat wil zeggen door consumptieniveaus (bijvoorbeeld van democratie). Dit betekent dat stated preferences data weinig valide zijn en daarom beter niet gebruikt kunnen worden in de studie van de economie van heterogene voorkeuren.

Deel I van dit proefschrift schuift de experienced preferences methode naar voren als de methode met het meeste potentieel voor het bestuderen van de economie van heterogene voorkeuren. Deel II van de thesis (Hoofdstuk 4 en 5) is vervolgens geheel gewijd aan het nader evalueren van deze methode. De experienced preferences methode is veelbelovend maar ook vrij nieuw en er is weinig bekend over de psychometrische kwaliteit van experienced preferences data. Pas als hier meer duidelijkheid over is, kunnen we een definitieve conclusie trekken omtrent de bruikbaarheid van deze methode.

De empirische input voor de experienced preferences methode zijn data over ervaren nut of subjectief welzijn. Deze data zijn gemeengoed in de psychologie maar worden relatief weinig gebruikt in de economie. Wel zijn ze sterk in opkomst. Hoofdstuk 4 geeft een uitgebreide introductie. Centrale vragen zijn welke indicatoren van subjectief welzijn er zoal zijn (bijvoorbeeld *geluk* en *tevredenheid*) en wat mogelijke toepassingen van deze indicatoren zijn, specifiek in beleidskringen.

Hoofdstuk 5 gaat uitgebreid in op de psychometrische kwaliteit, dat wil zeggen de validiteit en betrouwbaarheid, van zowel indicatoren van subjectief welzijn als de daarop gebaseerde maatstaven van verschillen in voorkeuren tussen groepen mensen. Om subjectief welzijn te meten, wordt vaak gebruik gemaakt van simpele vragen zoals “Hoe gelukkig bent u op een schaal van 1 tot 10?” of “Hoe tevreden bent u met uw leven als geheel?” Dit soort vragen blijkt zinvolle informatie op te leveren. De antwoorden van respondenten voorspellen het risico van zelfmoord en correleren bijvoorbeeld met fysiologische kenmerken zoals

bloeddruk. Daarnaast correleert het zelf ingeschatte welzijn met de inschatting van derden zoals de echtgeno(o)t(e) van een respondent of vrienden en familie. Eveneens is het zo dat mensen die zeggen gelukkig te zijn vaker glimlachen. Tegelijkertijd is het verstandig om er niet van uit te gaan dat als indicatoren van subjectief welzijn valide en betrouwbaar zijn, dit automatisch betekent dat experienced preferences data ook valide en betrouwbaar zijn. Daarom bevat Hoofdstuk 5 een uitgebreide empirische analyse van de psychometrische kwaliteit van experienced preferences maatstaven. We schatten ervaren nut als functie van specifieke factoren, afzonderlijk voor verschillende groepen. Heterogeniteit in deze nutsfuncties geeft de heterogeniteit in voorkeuren tussen de groepen aan. Door gebruik te maken van twee verschillende indicatoren van ervaren nut, te weten geluk en tevredenheid, krijgen we ook twee maatstaven van de heterogeniteit in voorkeuren. Dit maakt het mogelijk om te kijken of de twee maatstaven consistent zijn met elkaar. Indien beide maatstaven consistent zijn, is dit bewijs dat experienced preferences data valide en betrouwbaar zijn. De analyse laat zien dat de maatstaf gebaseerd op geluk inderdaad consistent is met de maatstaf gebaseerd op tevredenheid: beide maatstaven geven hetzelfde beeld van de heterogeniteit in voorkeuren tussen de groepen in de analyse. De conclusie is daarom dat de psychometrische kwaliteit van experienced preferences data voldoende is om deze te gebruiken in het bestuderen van de economie van heterogene voorkeuren.

De eerste twee delen van deze thesis vormen de voorbereiding voor het toepassen van de experienced preferences methode voor het feitelijk empirisch bestuderen van verschillen in voorkeuren in Deel III (Hoofdstuk 6 en 7). In de literatuur over heterogene voorkeuren krijgen twee kwesties prominente aandacht, namelijk een mogelijke Protestantse werkethiek en culturele verschillen tussen Oost- en West-Duitsers, in het bijzonder verschillen die zouden kunnen verklaren waarom Oost-Duitsland 20 jaar na de eenwording economisch nog altijd minder presteert dan West-Duitsland. Ruim een eeuw geleden sprak Max Weber over een Protestantse werkethiek. Hij legde een link tussen Calvinisme en de opkomst van kapitalistische systemen en instituties. Hij benadrukte daarbij de specifiek Protestantse visie die werk als iets positiefs ziet, een roeping. In navolging van Weber hebben veel wetenschappers empirisch onderzocht of Protestanten inderdaad een hogere werkethiek hebben. Vooralsnog hebben deze analyses echter niet tot eenduidig bewijs geleid. Een mogelijke verklaring is dat vrijwel al het onderzoek gebruik heeft gemaakt van stated preferences data met de daarbij behorende meetproblemen (zie boven). Hoofdstuk 6 kijkt naar het (negatieve) effect van werkloos zijn op het subjectief welzijn van mensen. Het belang van het hebben van een baan voor tevredenheid en geluk zegt iets over de houding van

mensen ten aanzien van werk. De resultaten laten zien dat Protestanten inderdaad meer waarde hechten aan werk: voor hen is de psychische pijn van werkloosheid significant hoger dan voor niet-Protestanten. Dit effect blijkt nog sterker te zijn op maatschappelijk niveau: in historisch Protestante samenlevingen heeft het niet hebben van een baan een groter negatief effect op welzijn dan in andere samenlevingen. Meer dan 100 jaar na Weber's originele studie vinden we in de resultaten van Hoofdstuk 6 dus sterke aanleiding om te concluderen dat er nog steeds een specifiek Protestantse werkethiek bestaat.

Hoofdstuk 7 analyseert de (mogelijke) culturele verschillen in voorkeuren tussen Oost- en West-Duitsers. Na de eenwording zijn formele instituties (e.g. wetgeving) gelijkgeschakeld tussen Oost- en West-Duitsland. Tegelijkertijd zijn er nog steeds flinke economische verschillen tussen beide voormalige Duitslanden. Aangezien deze verschillen niet verklaard kunnen worden uit verschillen in formele instituties, hebben onderzoekers zich toegelegd op andersoortige verklaringen, in het bijzonder op mogelijke culturele verschillen tussen de twee. De idee is dat 40 jaar Communisme een waardenerfenis heeft achtergelaten waardoor Oost-Duitsers minder goed aangepast zijn aan een kapitalistisch systeem; er zou sprake zijn van een *Mauer im Kopf*. De empirische strategie is om heterogene nutsfuncties te schatten voor Oost- en West-Duitsers. De gevonden verschillen ondersteunen de idee van een Mauer im Kopf als verklaring voor de economische achterstand van Oost-Duitsland niet. Zo lijken Oost-Duitsers niet minder maar juist meer waarde te hechten aan inkomen en aan de status van hun werk. Samen met Hoofdstuk 6 laat Hoofdstuk 7 zien dat de experienced preferences methode veel toegevoegde waarde heeft bij het bestuderen van de economie van heterogene voorkeuren.

Het afsluitende deel van de thesis, Deel IV, beslaat één hoofdstuk (Hoofdstuk 8). Dit hoofdstuk vat kort de belangrijkste bevindingen van het proefschrift samen en kijkt vooruit naar toekomstige ontwikkelingen in de studie van de economie van heterogene voorkeuren. Een belangrijke mogelijkheid voor toekomstig onderzoek is het toepassen van de experienced preferences methode in de ontwikkeling van een empirisch raamwerk van culturele verschillen tussen landen zoals eerder ontwikkeld door Geert Hofstede en anderen. Er is nog veel te leren over de opzichten waarin de voorkeuren van mensen aan elkaar gelijk zijn of juist van elkaar verschillen. De experienced preferences methode kan een nuttige bijdrage leveren aan het systematisch onderzoeken hiervan. Dit proefschrift heeft daartoe een belangrijke aanzet gegeven.

Curriculum Vitae

André van Hoorn was born in Helmond, the Netherlands. After obtaining a degree in applied physics from Eindhoven Polytechnic, he studied Economics in Tilburg, graduating in 2003. In 2005 he completed the postgraduate program Financial and Economic Policy (BoFEB) from Erasmus University Rotterdam (in combination with several Dutch ministries and DNB Central Bank). As part of this program André worked at the Research, Strategy & International Affairs unit of the Ministry of Economic Affairs.

André started working on his PhD research at the Department of Economics of the Nijmegen School of Management in 2005. Between 2006 and 2009 André also was a lecturer at the same institute, teaching, coordinating and developing courses in Micro- & Macroeconomics and in Behavioral & Experimental Economics. During his PhD project, André spent six months as a trainee at the Organisation for Economic Co-operation and Development (OECD), where he worked on the global project Measuring the Progress of Societies, part of the World Forum on Statistics, Knowledge and Policy. He has also been a visiting researcher at Erasmus University Rotterdam.

André's research interests include culture and (informal) institutions, happiness and welfare, and behavioral economics. His work has been published or is forthcoming in a variety of internationally recognized outlets, including the *Journal of International Business Studies*, the *Journal of Economic Behavior & Organization*, and *Public Choice*. In addition, he has presented his work at various international conferences and workshops. He has also contributed several pieces to general-interest magazines.

As of 2010 André is associated with the Faculty of Economics & Business of the University of Groningen as an assistant professor. At Groningen he teaches both at the graduate and the undergraduate level, with a focus on comparative studies.